```
RESULT 20
MMIIGA/c
LOCUS        MMIIGA             585 bp    DNA        linear    ROD 24-FEB-1999
DEFINITION   Mouse Ii gene for Ia antigen associated invariant chains Ii31/Ii41
             and Ii41) exon 1.
ACCESSION    X05428
VERSION      X05428.1  GI:52626
KEYWORDS     alternative splicing; antigen; class II antigen; glycoprotein; Ia
             antigen; Ia antigen-associated invariant chain; repetitive
             sequence.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 585)
  AUTHORS    Koch,N., Lauer,W., Habicht,J. and Dobberstein,B.
  TITLE      Primary structure of the gene for the murine Ia antigen-associated
             invariant chains (Ii). An alternatively spliced exon encodes a
             cysteine-rich domain highly homologous to a repetitive sequence of
             thyroglobulin
  JOURNAL    EMBO J. 6 (6), 1677-1683 (1987)
  MEDLINE    87275861
   PUBMED    3038530
COMMENT      see x05429-30 for remaining exons.
FEATURES             Location/Qualifiers
     source          1..585
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="AKR"
                     /db_xref="taxon:10090"
     misc_feature    22..36
                     /note="conserved class II MHC antigen 15-mer sequence"
     promoter        166..170
                     /note="pot. CAAT-box"
     misc_feature    178..183
                     /note="SpI protein binding site"
     promoter        226..230
                     /note="pot. TATA-box"
     misc_feature    230
                     /note="put. treanscription start site"
     repeat_region   237..312
                     /note="repetitive CAG sequence"
     mRNA            250..>419
     exon            250..>419
                     /number=1
     CDS             343..>419
                     /note="unnamed protein product; Ii protein"
                     /codon_start=1
                     /protein_id="CAA29010.1"
                     /db_xref="GI:52627"
                     /db_xref="GOA:P04441"
                     /db_xref="SWISS-PROT:P04441"
                     /translation="MDDQRDLISNHEQLPILGNRPRPE"
     intron          420..>585
                     /number=1
ORIGIN

Query Match            100.0%;  Score 18;  DB 10;  Length 585;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db    354 TTGGTCATCCATGGCTCT 337
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 14:47:42 ; Search time 1543 Seconds
                                          (without alignments)
                                          505.622 Million cell updates/sec

Title: US-10-054-387-40
Perfect score: 18
Sequence: 1 ttggtcatccatggctct 18

Scoring table: IDENTITY NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 500 summaries

Database :   GenEmbl:*
            1:   gb_ba:*
            2:   gb_htg:*
            3:   gb_in:*
            4:   gb_om:*
            5:   gb_ov:*
            6:   gb_pat:*
            7:   gb_ph:*
            8:   gb_pl:*
            9:   gb_pr:*
           10:   gb_ro:*
           11:   gb_sts:*
           12:   gb_sy:*
           13:   gb_un:*
           14:   gb_vi:*
           15:   em_ba:*
           16:   em_fun:*
           17:   em_hum:*
           18:   em_in:*
           19:   em_mu:*
           20:   em_om:*
           21:   em_or:*
           22:   em_ov:*
           23:   em_pat:*
           24:   em_ph:*
           25:   em_pl:*
           26:   em_ro:*
           27:   em_sts:*
           28:   em_un:*
           29:   em_vi:*
           30:   em_htg_hum:*
           31:   em_htg_inv:*
           32:   em_htg_other:*
           33:   em_htg_mus:*
           34:   em_htg_pln:*
           35:   em_htg_rod:*
           36:   em_htg_man:*
           37:   em_htg_vrt:*
           38:   em_sy:*
           39:   em_htgo_hum:*
           40:   em_htgo_mus:*
           41:   em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	6	BD274814	BD274814 CANCER CE
2		18	100.0	18	6	AR205280	AR205280 Sequence
3		18	100.0	24	6	I91647	I91647 Sequence 1
4	c	18	100.0	103	6	BD274839	BD274839 CANCER CE
5	c	18	100.0	103	6	AR205305	AR205305 Sequence
6	c	18	100.0	107	6	BD274846	BD274846 CANCER CE
7	c	18	100.0	107	6	AR205312	AR205312 Sequence
8	c	18	100.0	124	6	BD274850	BD274850 CANCER CE
9	c	18	100.0	124	6	AR205316	AR205316 Sequence
10	c	18	100.0	141	6	I91650	I91650 Sequence 4
11	c	18	100.0	148	6	BD274849	BD274849 CANCER CE
12	c	18	100.0	148	6	AR205315	AR205315 Sequence
13	c	18	100.0	169	6	BD274843	BD274843 CANCER CE
14	c	18	100.0	169	6	AR205309	AR205309 Sequence
15	c	18	100.0	190	6	BD274848	BD274848 CANCER CE
16	c	18	100.0	190	6	AR205314	AR205314 Sequence
17	c	18	100.0	252	10	MMIIR	X07129 Mouse mIi m
18	c	18	100.0	487	11	BV096208	BV096208 RPAMMSEQ0
19	c	18	100.0	508	10	MUSIFNRES	M35872 Mouse MHC c
20	c	18	100.0	585	10	MMIIGA	X05428 Mouse Ii ge
21	c	18	100.0	1375	10	BC003476	BC003476 Mus muscu
22	c	18	100.0	1576	10	BC061489	BC061489 Mus muscu
23	c	18	100.0	2580	6	AR199681	AR199681 Sequence
24	c	18	100.0	9658	2	MMIIG	X13414 Murine I ge
25	c	18	100.0	203664	2	AC139759	AC139759 Mus muscu
26		17	94.4	18	6	BD274827	BD274827 CANCER CE
27		17	94.4	18	6	BD274828	BD274828 CANCER CE
28		17	94.4	18	6	AR205293	AR205293 Sequence
29		17	94.4	18	6	AR205294	AR205294 Sequence
30	c	17	94.4	146937	2	BX890568	BX890568 Danio rer
31		17	94.4	183607	2	AC121176	AC121176 Rattus no
32		17	94.4	231571	2	AC126656	AC126656 Rattus no
33	c	17	94.4	239467	2	AC111384	AC111384 Rattus no
34		16.4	91.1	521	6	AX795517	AX795517 Sequence
35		16.4	91.1	521	6	AX796809	AX796809 Sequence
36		16.4	91.1	521	6	AX797071	AX797071 Sequence
37		16.4	91.1	42433	10	AL603822	AL603822 Mouse DNA
38	c	16.4	91.1	79271	2	AC023361	AC023361 Mus muscu
39	c	16.4	91.1	102448	10	AL928696	AL928696 Mouse DNA
40		16.4	91.1	113950	2	AC114395	AC114395 Zea mays
41		16.4	91.1	113950	2	AC114395	AC114395 Zea mays
42		16.4	91.1	126323	8	AC132215	AC132215 Genomic s
43		16.4	91.1	134514	2	AC025192	AC025192 Homo sapi
44		16.4	91.1	146306	2	AF248716	AF248716 Mus muscu
45		16.4	91.1	151820	2	AC140494	AC140494 Homo sapi
46		16.4	91.1	153988	2	AC138308	AC138308 Mus muscu
47		16.4	91.1	157081	2	AC023084	AC023084 Homo sapi
48	c	16.4	91.1	157324	9	AC013643	AC013643 Homo sapi
49		16.4	91.1	158022	9	AL445306	AL445306 Human DNA
50	c	16.4	91.1	158230	1	AF440523	AF440523 Pseudomon
51		16.4	91.1	164746	9	AL732431	AL732431 Human DNA
52		16.4	91.1	166299	9	AC092602	AC092602 Homo sapi
53		16.4	91.1	166971	10	AL845529	AL845529 Mouse DNA
54		16.4	91.1	175380	9	AF235104	AF235104 Homo sapi
55		16.4	91.1	177059	2	AL954846	AL954846 Danio rer
56		16.4	91.1	178508	2	AC023040	AC023040 Homo sapi
57		16.4	91.1	178563	2	AP002419	AP002419 Homo sapi
58		16.4	91.1	181342	2	AC023556	AC023556 Homo sapi
59		16.4	91.1	189983	2	AC133524	AC133524 Mus muscu
60		16.4	91.1	191540	9	AC010984	AC010984 Homo sapi
61		16.4	91.1	196501	9	AC005908	AC005908 Homo sapi
62		16.4	91.1	202373	2	AC122090	AC122090 Rattus no
63	c	16.4	91.1	203811	2	AC140481	AC140481 Homo sapi
64		16.4	91.1	208279	5	AL929009	AL929009 Zebrafish
65		16.4	91.1	216259	2	AC099261	AC099261 Rattus no

#	c			position	n	accession	description
66		16.4	91.1	216975	2	AC133376	AC133376 Rattus no
67		16.4	91.1	221397	2	AC098768	AC098768 Rattus no
68		16.4	91.1	225098	2	AC126163	AC126163 Rattus no
69		16.4	91.1	229045	9	AC008680	AC008680 Homo sapi
70		16.4	91.1	230372	2	AC073693	AC073693 Mus muscu
71		16.4	91.1	234429	9	AF311103	AF311103 Homo sapi
72		16.4	91.1	238657	10	AC116592	AC116592 Mus muscu
73		16.4	91.1	242279	2	AC097300	AC097300 Rattus no
74		16.4	91.1	247554	2	AC134082	AC134082 Rattus no
75		16.4	91.1	249121	2	AC096843	AC096843 Rattus no
76		16.4	91.1	252458	2	AC094729	AC094729 Rattus no
77	c	16.4	91.1	252473	2	AC109671	AC109671 Rattus no
78		16.4	91.1	258222	2	AC103417	AC103417 Rattus no
79		16.4	91.1	272657	2	AC113791	AC113791 Rattus no
80		16.4	91.1	292420	2	AC020873	AC020873 Mus muscu
81		16.4	91.1	292513	2	AC123088	AC123088 Rattus no
82		16	88.9	349989	6	AX647177	AX647177 Sequence
83		16	88.9	18	6	BD274826	BD274826 CANCER CE
84		16	88.9	18	6	BD274829	BD274829 CANCER CE
85		16	88.9	18	6	AR205292	AR205292 Sequence
86		16	88.9	18	6	AR205295	AR205295 Sequence
87		16	88.9	2734	8	MCU84889	U84889 Mesembryant
88		16	88.9	31248	6	AX707115	AX707115 Sequence
89		16	88.9	31248	6	AF235050	AF235050 Streptomy
90		16	88.9	35359	6	AX707116	AX707116 Sequence
91		16	88.9	116671	9	AC087264	AC087264 Pan trogl
92		16	88.9	126559	9	AC087253	AC087253 Pan trogl
93		16	88.9	131811	2	BX842677	BX842677 Danio rer
94		16	88.9	147363	8	AP003487	AP003487 Oryza sat
95		16	88.9	153602	8	AL589949	AL589949 Human DNA
96	c	16	88.9	154180	8	AP000399	AP000399 Oryza sat
97	c	16	88.9	160314	2	BX890552	BX890552 Danio rer
98	c	16	88.9	170587	10	AC110255	AC110255 Mus muscu
99	c	16	88.9	171571	9	AC142342	AC142342 Pan trogl
100	c	16	88.9	171905	9	AC002080	AC002080 Human BAC
101	c	16	88.9	174393	9	AC123232	AC123232 Rattus no
102	c	16	88.9	178567	2	BX571836	BX571836 Danio rer
103		16	88.9	185847	9	AC121334	AC121334 Homo sapi
104		16	88.9	202257	2	AC119314	AC119314 Rattus no
105		16	88.9	211409	2	AC127104	AC127104 Danio rer
106		16	88.9	226326	2	AL929507	AL929507 Danio rer
107		16	88.9	229332	2	BX284665	BX284665 Danio rer
108		16	88.9	235460	2	AC107798	AC107798 Mus muscu
109		16	88.9	239084	5	AL845437	AL845437 Zebrafish
110	c	16	88.9	240454	2	BX510324	BX510324 Danio rer
111		16	88.9	241990	2	AC123449	AC123449 Rattus no
112		16	88.9	251132	2	AC098162	AC098162 Rattus no
113	c	16	88.9	262782	2	AC115230	AC115230 Rattus no
114		16	88.9	279534	2	AC114035	AC114035 Rattus no
115	c	15.4	85.6	381	8	AF483856	AF483856 Phaseolus
116		15.4	85.6	536	2	AX385329	AX385329 Sequence
117	c	15.4	85.6	587	11	G39221	G39221 Z20095 Zebr
118		15.4	85.6	836	8	HSA322886	AJ322886 Homo sapi
119	c	15.4	85.6	1150	6	RNINVG34	AX827833 Sequence
120	c	15.4	85.6	1150	10	BC059152	X13044 Rat mRNA fo
121	c	15.4	85.6	1209	10	BC059152	BC059152 Rattus no
122		15.4	85.6	75454	2	AR179661	AR179661 Sequence
123	c	15.4	85.6	1353	6	AX827837	AX827837 Sequence
124	c	15.4	85.6	1353	10	RNMEC2I	X14254 Rat mRNA fo
125	c	15.4	85.6	381	8	AC142404	AC142404 Homo sapi
126		15.4	85.6	38820	2	AC146474	AC146474 Hylobates
127	c	15.4	85.6	41735	9	AC134050	AC134050 Homo sapi
128		15.4	85.6	43725	2	AC142403	AC142403 Homo sapi
129	c	15.4	85.6	72666	9	AC093614	AC093614 Homo sapi
130		15.4	85.6	73000	8	AP004134	AP004134 Oryza sat
131	c	15.4	85.6	75454	2	AC011433	AC011433 Homo sapi
132	c	15.4	85.6	76581	9	AL355304	AL355304 Human DNA
133	c	15.4	85.6	80659	9	AC004553	AC004553 Homo sapi
134		15.4	85.6	88876	2	AC147595	AC147595 Mus muscu
135	c	15.4	85.6	98377	8	AP003830	AP003830 Oryza sat
136		15.4	85.6	98377	8	AP003830	AP003830 Oryza sat
137		15.4	85.6	98980	2	AP004215	AP004215 Oryza sat
138	c	15.4	85.6	100000	9	AP000496	AP000496 Homo sapi
139		15.4	85.6	103851	2	AP005616	AP005616 Oryza sat
140		15.4	85.6	110000	2	AC145908_1	Continuation (2 of
141		15.4	85.6	110000	2	AC145908_2	Continuation (3 of
142		15.4	85.6	110000	2	BX572080_1	Continuation (2 of
143		15.4	85.6	110250	9	AP005295	AP005295 Oryza sat
144	c	15.4	85.6	112031	9	AL445487	AL445487 Human DNA
145		15.4	85.6	114501	2	AP005093	AP005093 Oryza sat
146		15.4	85.6	117226	8	AP006626	AP006626 Oryza sat
147		15.4	85.6	119329	8	AC097368	AC097368 Oryza sat
148	c	15.4	85.6	120528	9	AL158817	AL158817 Human DNA
149		15.4	85.6	123057	2	BX890549	BX890549 Danio rer
150	c	15.4	85.6	124138	8	AC093017	AC093017 Oryza sat
151		15.4	85.6	125630	2	AC090650	AC090650 Arabidops
152		15.4	85.6	126734	8	AP003849	AP003849 Oryza sat
153	c	15.4	85.6	131400	2	AC117642	AC117642 Mus muscu
154		15.4	85.6	134059	2	AC016284	AC016284 Homo sapi
155	c	15.4	85.6	135064	2	AL627255	AL627255 Danio rer
156		15.4	85.6	137525	8	AP005397	AP005397 Oryza sat
157	c	15.4	85.6	138262	2	AC073864	AC073864 Homo sapi
158		15.4	85.6	141791	9	AP005582	AP005582 Oryza sat
159		15.4	85.6	142525	5	AC105752	AC105752 Homo sapi
160	c	15.4	85.6	142859	2	AC008663	AC008663 Homo sapi
161		15.4	85.6	149260	2	AC116786	AC116786 Mus muscu
162	c	15.4	85.6	150972	2	AL512380	AL512380 Human DNA
163		15.4	85.6	152019	2	BX663494	BX663494 Danio rer
164	c	15.4	85.6	153777	2	AP005543	AP005543 Oryza sat
165		15.4	85.6	155526	2	AC013371	AC013371 Homo sapi
166		15.4	85.6	157241	2	AC046194	AC046194 Homo sapi
167		15.4	85.6	157493	2	AC027068	AC027068 Homo sapi
168		15.4	85.6	157682	5	AC110560	AC110560 Mus muscu
169	c	15.4	85.6	157816	2	AC093666	AC093666 Mus muscu
170		15.4	85.6	157907	2	BX510302	BX510302 Danio rer
171	c	15.4	85.6	158139	5	BX120016	BX120016 Danio rer
172	c	15.4	85.6	158901	2	AC145786	AC145786 Silurana
173	c	15.4	85.6	159401	9	AC104671	AC104671 Mus muscu
174		15.4	85.6	159523	2	BX321908	BX321908 Danio rer
175	c	15.4	85.6	161652	9	AC119346	AC119346 Rattus no
176		15.4	85.6	162820	9	AL138765	AL138765 Human DNA
177		15.4	85.6	163429	2	AC114473	AC114473 Mus muscu
178	c	15.4	85.6	163621	2	AL445704	AL445704 Human DNA
179		15.4	85.6	164197	2	AC140844	AC140844 Mus muscu
180		15.4	85.6	164823	2	AC009567	AC009567 Homo sapi
181		15.4	85.6	165902	9	AC106763	AC106763 Homo sapi
182		15.4	85.6	167380	9	AP006242	AP006242 Homo sapi
183		15.4	85.6	167456	2	CNS09SBY	BX649219 Oryza sat
184		15.4	85.6	167694	2	AC055737	AC055737 Oryza sat
185		15.4	85.6	169709	9	AC104183	AC104183 Homo sapi
186		15.4	85.6	170249	2	AC068799	AC068799 Homo sapi
187		15.4	85.6	170849	2	AC109803	AC109803 Homo sapi
188		15.4	85.6	171087	2	AC011987	AC011987 Homo sapi
189		15.4	85.6	171701	2	AC090075	AC090075 Homo sapi
190		15.4	85.6	177938	9	AC103670	AC103670 Mus muscu
191		15.4	85.6	178372	2	BX569797	BX569797 Danio rer
192	c	15.4	85.6	179269	9	AC093627	AC093627 Homo sapi
193	c	15.4	85.6	179414	9	AC098933	AC098933 Homo sapi
194		15.4	85.6	181567	5	AL354811	AL354811 Human DNA
195		15.4	85.6	182141	5	AL954645	AL954645 Zebrafish
196		15.4	85.6	182533	2	BX276188	BX276188 Danio rer
197		15.4	85.6	183009	9	AC116350	AC116350 Homo sapi
198	c	15.4	85.6	183491	2	BX539325	BX539325 Danio rer
199		15.4	85.6	183754	2	AC023183	AC023183 Homo sapi
200		15.4	85.6	184908	10	AC121981	AC121981 Mus muscu
201		15.4	85.6	186053	2	AC140318	AC140318 Homo sapi
202		15.4	85.6	186651	2	AC146680	AC146680 Otolemur
203		15.4	85.6	186685	9	AC116676	AC116676 Mus muscu
204	c	15.4	85.6	187270	9	BX001029	BX001029 Danio rer
205		15.4	85.6	188992	9	AC114224	AC114224 Rattus no
206		15.4	85.6	189854	2	AC093686	AC093686 Homo sapi
207	c	15.4	85.6	191263	2	BX571712	BX571712 Danio rer
208	c	15.4	85.6	192001	10	AC068376	AC068376 Homo sapi
209	c	15.4	85.6	192016	2	AC102426	AC102426 Mus muscu
210		15.4	85.6	192344	2	AC119667	AC119667 Rattus no
211	c	15.4	85.6	193255	10	AC127349	AC127349 Mus muscu

#							Description
c 212	15.4	85.6	193277	10	AC104324	AC104324	Mus muscu
213	15.4	85.6	193298	10	BX088708	BX088708	Danio rer
214	15.4	85.6	193458	10	AL645606	AL645606	Mouse DNA
c 215	15.4	85.6	194420	2	AC124133	AC124133	Mus muscu
216	15.4	85.6	194612	2	AC021118	AC021118	Homo sapi
217	15.4	85.6	195383	2	AC126472	AC126472	Homo sapi
c 218	15.4	85.6	195776	5	BX088566	BX088566	Zebrafish
219	15.4	85.6	195932	2	AC021710	AC021710	Homo sapi
c 220	15.4	85.6	196027	2	BX897690	BX897690	Danio rer
221	15.4	85.6	196492	2	AC139725	AC139725	Danio rer
222	15.4	85.6	196641	9	AC093415	AC093415	Homo sapi
c 223	15.4	85.6	198727	2	AC122401	AC122401	Mus muscu
224	15.4	85.6	199517	9	AL353741	AL353741	Human DNA
c 225	15.4	85.6	200074	2	BX537313	BX537313	Danio rer
226	15.4	85.6	200252	2	AC133192	AC133192	Mus muscu
227	15.4	85.6	200586	2	AC119941	AC119941	Mus muscu
c 228	15.4	85.6	200697	5	AL935198	AL935198	Zebrafish
229	15.4	85.6	200858	8	AP005583	AP005583	Oryza sat
230	15.4	85.6	201170	2	AC022259	AC022259	Homo sapi
c 231	15.4	85.6	201220	8	AC146868	AC146868	Silurana
c 232	15.4	85.6	205806	2	AC107836	AC107836	Mus muscu
233	15.4	85.6	206302	2	AC133125	AC133125	Homo sapi
234	15.4	85.6	208763	2	AC123366	AC123366	Rattus no
c 235	15.4	85.6	210005	2	BX649516	BX649516	Danio rer
236	15.4	85.6	210752	2	AC098606	AC098606	Rattus no
237	15.4	85.6	210982	2	BX539345	BX539345	Danio rer
238	15.4	85.6	216109	10	AC091777	AC091777	Mus muscu
c 239	15.4	85.6	216272	2	AC092712	AC092712	Mus muscu
240	15.4	85.6	217871	2	BX284621	BX284621	Danio rer
241	15.4	85.6	219346	10	AC098740	AC098740	Mus muscu
c 242	15.4	85.6	219550	2	AC135904	AC135904	Rattus no
243	15.4	85.6	220992	2	AC141153	AC141153	Rattus no
c 244	15.4	85.6	221189	2	AC096341	AC096341	Rattus no
245	15.4	85.6	222105	2	BX119971	BX119971	Danio rer
246	15.4	85.6	222163	10	AL731558	AL731558	Mouse DNA
c 247	15.4	85.6	222855	2	AC096703	AC096703	Rattus no
c 248	15.4	85.6	223000	5	BX004839	BX004839	Zebrafish
249	15.4	85.6	224584	5	BX470160	BX470160	Zebrafish
c 250	15.4	85.6	225119	2	AC097781	AC097781	Rattus no
251	15.4	85.6	226450	2	BX088714	BX088714	Danio rer
c 252	15.4	85.6	226642	2	AC103918	AC103918	Rattus no
253	15.4	85.6	227550	2	AC102686	AC102686	Mus muscu
c 254	15.4	85.6	229403	2	AC115974	AC115974	Mus muscu
255	15.4	85.6	229823	2	AC145199	AC145199	Mus muscu
256	15.4	85.6	230030	2	AC130854	AC130854	Rattus no
c 257	15.4	85.6	231698	2	AC118117	AC118117	Rattus no
258	15.4	85.6	231937	2	AC129153	AC129153	Rattus no
c 259	15.4	85.6	232555	2	AC128470	AC128470	Rattus no
c 260	15.4	85.6	232720	10	AC124314	AC124314	Mus muscu
261	15.4	85.6	233296	2	AC118193	AC118193	Mus muscu
c 262	15.4	85.6	233841	5	BX640576	BX640576	Danio rer
263	15.4	85.6	234065	2	AC140765	AC140765	Rattus no
264	15.4	85.6	234592	2	AC124662	AC124662	Mus muscu
c 265	15.4	85.6	236450	10	AC118545	AC118545	Mus muscu
266	15.4	85.6	237200	2	AC097766	AC097766	Rattus no
267	15.4	85.6	238337	2	AC127777	AC127777	Rattus no
268	15.4	85.6	238353	2	AC125551	AC125551	Rattus no
269	15.4	85.6	239196	2	AC114045	AC114045	Rattus no
270	15.4	85.6	239237	2	AC094621	AC094621	Rattus no
271	15.4	85.6	239948	2	AC129116	AC129116	Rattus no
272	15.4	85.6	241289	2	AC124662	AC124662	Mus muscu
273	15.4	85.6	242905	2	AC135015	AC135015	Mus muscu
c 274	15.4	85.6	243259	2	AC127409	AC127409	Rattus no
275	15.4	85.6	244786	2	AC125943	AC125943	Rattus no
276	15.4	85.6	245071	2	AC145961	AC145961	Gallus ga
277	15.4	85.6	250080	2	AC095289	AC095289	Rattus no
278	15.4	85.6	250386	2	AC096198	AC096198	Rattus no
c 279	15.4	85.6	250512	2	AC095618	AC095618	Rattus no
c 280	15.4	85.6	253885	2	AC094257	AC094257	Rattus no
c 281	15.4	85.6	255275	2	AC147444	AC147444	Gallus ga
282	15.4	85.6	255563	2	AC126930	AC126930	Mus muscu
283	15.4	85.6	261278	2	AC130742	AC130742	Rattus no
284	15.4	85.6	261498	2	AC073823	AC073823	Mus muscu

#							Description
c 285	15.4	85.6	261659	2	AC109683	AC109683	Rattus no
286	15.4	85.6	261927	2	AC126891	AC126891	Rattus no
287	15.4	85.6	266391	2	AC097092	AC097092	Rattus no
c 288	15.4	85.6	267978	2	AC099444	AC099444	Rattus no
289	15.4	85.6	268213	2	AC096961	AC096961	Rattus no
290	15.4	85.6	272266	2	AC096358	AC096358	Rattus no
291	15.4	85.6	273544	2	AC099419	AC099419	Rattus no
292	15.4	85.6	274688	2	AC127210	AC127210	Rattus no
293	15.4	85.6	277191	2	AC132658	AC132658	Rattus no
294	15.4	85.6	278954	2	AC121436	AC121436	Rattus no
295	15.4	85.6	279725	2	AC131174	AC131174	Rattus no
296	15.4	85.6	280810	2	AC046188	AC046188	Homo sapi
c 297	15.4	85.6	338387	2	AC016150	AC016150	Homo sapi
298	15	83.3	18	6	BD274813	BD274813	CANCER CE
299	15	83.3	18	6	BD274815	BD274815	CANCER CE
300	15	83.3	18	6	AR205279	AR205279	Sequence
301	15	83.3	18	6	AR205281	AR205281	Sequence
c 302	15	83.3	300	11	G28826	G28826	D11S1067 cS
c 303	15	83.3	541	11	BV062152	BV062152	S20BP6272
304	15	83.3	613	6	AR225551	AR225551	Sequence
305	15	83.3	613	6	AX321707	AX321707	Sequence
c 306	15	83.3	629	11	BV067938	BV067938	S209P6035
307	15	83.3	856	6	BD018337	BD018337	Novel gen
c 308	15	83.3	856	6	BD098275	BD098275	Novel gen
309	15	83.3	1484	8	X76532	TAPWR5PI	Triticum ae
c 310	15	83.3	1490	8	X15869	HVNADPHP	Barley mRNA
311	15	83.3	2891	2	AK025237	AK025237	Homo sapi
312	15	83.3	5085	1	AF088132	AF088132	Herbaspir
c 313	15	83.3	6476	6	AX321897	AX321897	Sequence
314	15	83.3	6477	6	AB018255	AB018255	Homo sapi
c 315	15	83.3	6480	9	AY194287	AY194287	Homo sapi
c 316	15	83.3	6871	9	BC054488	BC054488	Homo sapi
c 317	15	83.3	8271	9	AL833062	HSM804373	Homo sapi
c 318	15	83.3	37664	2	AC100940	AC100940	Mus muscu
c 319	15	83.3	38791	2	AC005597	AC005597	Homo sapi
c 320	15	83.3	44251	3	AC084599	CBRG42P09	Caenorhab
321	15	83.3	61566	9	AL139346	AL139346	Human DNA
322	15	83.3	70355	10	Continuation (4 of	AC113764_3	
c 323	15	83.3	103260	2	AL713995	AL713995	Mouse DNA
324	15	83.3	110000	2	Continuation (3 of	AC113764_2	
325	15	83.3	110801	2	AL390054	AL390054	Homo sapi
c 326	15	83.3	118696	2	AC108843	AC108843	Mus muscu
327	15	83.3	118696	5	AC108843	AC108843	Mus muscu
328	15	83.3	128648	5	AC026025	AC026025	Homo sapi
c 329	15	83.3	128639	5	BX548009	BX548009	Zebrafish
330	15	83.3	135431	2	AC103393	AC103393	Mus muscu
331	15	83.3	136643	2	AC143470	AC143470	Macaca mu
c 332	15	83.3	136969	5	BX663493	BX663493	Danio rer
333	15	83.3	139224	9	AC093857	AC093857	Homo sapi
c 334	15	83.3	140597	9	AC074270	AC074270	Homo sapi
335	15	83.3	140080	9	AC013755	AC013755	Homo sapi
c 336	15	83.3	146216	9	AL161892	AL161892	Human DNA
337	15	83.3	146850	5	AC141349	AC141349	Rattus no
c 338	15	83.3	148322	5	BX072568	BX072568	Zebrafish
339	15	83.3	149237	5	BX284700	BX284700	Danio rer
340	15	83.3	152295	5	BX005103	BX005103	Zebrafish
341	15	83.3	152946	10	AC124021	AC124021	Mus muscu
c 342	15	83.3	154604	5	AL954739	AL954739	Zebrafish
343	15	83.3	155469	10	AC124020	AC124020	Mus muscu
c 344	15	83.3	156000	8	OSJNOO050	OSJNOO050	Oryza sat
345	15	83.3	157691	5	AL935143	AL935143	Zebrafish
346	15	83.3	159034	5	BX470154	BX470154	Danio rer
347	15	83.3	159817	5	BX322798	BX322798	Danio rer
348	15	83.3	161218	5	AL772148	AL772148	Zebrafish
349	15	83.3	161812	5	BX005045	BX005045	Zebrafish
350	15	83.3	162017	5	AC141876	AC141876	Mus muscu
351	15	83.3	162029	5	BX323854	BX323854	Danio rer
352	15	83.3	162215	9	AC004896	AC004896	Homo sapi
353	15	83.3	163098	10	AC124022	AC124022	Mus muscu
c 354	15	83.3	163348	5	BX842613	BX842613	Danio rer
355	15	83.3	165085	9	AC073607	AC073607	Homo sapi
c 356	15	83.3	167188	2	AC102497	AC102497	Mus muscu
357	15	83.3	169479	2	BX890540	BX890540	Danio rer

No	c	Accession	Length	%	Score	n	Description
358		AC020907	169891	83.3	15	9	AC020907 Homo sapi
359		AC011045	171637	83.3	15	2	AC011045 Mus muscu
360		AC024682	174875	83.3	15	2	AC024682 Homo sapi
361		AL808110	175963	83.3	15	10	AL808110 Mouse DNA
362	c	BX322785	176361	83.3	15	2	BX322785 Danio rer
363		BX897691	176626	83.3	15	2	BX897691 Danio rer
364		AP000751	176713	83.3	15	9	AP000751 Homo sapi
365	c	AC124026	178243	83.3	15	2	AC124026 Mus muscu
366		BX004847	179903	83.3	15	2	BX004847 Danio rer
367		AC125819	180035	83.3	15	2	AC125819 Rattus no
368		BX000473	180366	83.3	15	5	BX000473 Zebrafish
369	c	BX537290	181265	83.3	15	2	BX537290 Danio rer
370		BX548007	182521	83.3	15	2	BX548007 Danio rer
371	c	BX323801	182831	83.3	15	2	BX323801 Danio rer
372	c	BX005453	185533	83.3	15	2	BX005453 Danio rer
373		AL954147	188067	83.3	15	5	AL954147 Zebrafish
374	c	AC023532	189412	83.3	15	2	AC023532 Homo sapi
375		AC140370	189786	83.3	15	10	AC140370 Mus muscu
376		AC119770	190382	83.3	15	2	AC119770 Rattus no
377		AC133619	191880	83.3	15	2	AC133619 Rattus no
378		AP001265	192913	83.3	15	9	AP001265 Homo sapi
379	c	AC026767	193167	83.3	15	10	AC026767 Mus muscu
380		AC069535	193795	83.3	15	2	AC069535 Mus muscu
381		AC096534	195270	83.3	15	9	AC096534 Homo sapi
382		BX088722	195359	83.3	15	5	BX088722 Zebrafish
383		AL935148	195396	83.3	15	5	AL935148 Zebrafish
384		AL928885	196857	83.3	15	5	AL928885 Zebrafish
385		BX571830	196976	83.3	15	2	BX571830 Danio rer
386	c	AC116840	197940	83.3	15	2	AC116840 Mus muscu
387		AL954343	198352	83.3	15	5	AL954343 Zebrafish
388		BX530058	199541	83.3	15	2	BX530058 Danio rer
389		AC112090	200030	83.3	15	2	AC112090 Rattus no
390		BX510311	200050	83.3	15	2	BX510311 Danio rer
391	c	AC125467	200304	83.3	15	10	AC125467 Mus muscu
392		BX511242	201584	83.3	15	2	BX511242 Danio rer
393		AC128271	201928	83.3	15	2	AC128271 Rattus no
394		AC128076	202866	83.3	15	2	AC128076 Rattus no
395		BX649245	203971	83.3	15	2	BX649245 Danio rer
396		AP002985	205152	83.3	15	9	AP002985 Homo sapi
397		AL591032	205573	83.3	15	10	AL591032 Mouse DNA
398		AL805965	205790	83.3	15	10	AL805965 Mouse DNA
399		AL732501	206856	83.3	15	10	AL732501 Mouse DNA
400	c	BX572103	207783	83.3	15	2	BX572103 Danio rer
401		BX005398	209859	83.3	15	5	BX005398 Zebrafish
402		BX119315	210930	83.3	15	5	BX119315 Zebrafish
403		AC136126	212221	83.3	15	2	AC136126 Rattus no
404		BX569782	214481	83.3	15	2	BX569782 Danio rer
405		AP001447	214701	83.3	15	9	AP001447 Homo sapi
406		AC109714	217088	83.3	15	2	AC109714 Rattus no
407		AC111489	217558	83.3	15	2	AC111489 Rattus no
408		BX322583	218143	83.3	15	2	BX322583 Mus muscu
409		BX470118	218913	83.3	15	2	BX470118 Danio rer
410		AC105644	220742	83.3	15	2	AC105644 Rattus no
411		AC106938	221323	83.3	15	2	AC106938 Rattus no
412		AL626777	221428	83.3	15	2	AL626777 Mus muscu
413		BX510345	222746	83.3	15	2	BX510345 Danio rer
414		AL672102	225611	83.3	15	2	AL672102 Mus muscu
415	c	AC073599	226256	83.3	15	10	AC073599 Mus muscu
416		AC098160	226328	83.3	15	2	AC098160 Rattus no
417		BX511306	226398	83.3	15	2	BX511306 Danio rer
418		AL805947	226542	83.3	15	5	AL805947 Zebrafish
419		AC095653	227788	83.3	15	2	AC095653 Rattus no
420		AC125885	227837	83.3	15	2	AC125885 Rattus no
421		BX119989	228548	83.3	15	2	BX119989 Danio rer
422		AC097668	230799	83.3	15	2	AC097668 Rattus no
423		AC128144	232042	83.3	15	2	AC128144 Rattus no
424		AC111679	232742	83.3	15	2	AC111679 Rattus no
425		BX293997	234419	83.3	15	2	BX293997 Danio rer
426		BX572104	239132	83.3	15	2	BX572104 Danio rer
427		AC131219	240446	83.3	15	2	AC131219 Rattus no
428	c	AC019279	241392	83.3	15	2	AC019279 Homo sapi
429		AC095256	242364	83.3	15	2	AC095256 Rattus no
430		BX571714	242372	83.3	15	2	BX571714 Danio rer
431	c	AC098749	244292	83.3	15	2	AC098749 Rattus no
432	c	AC125965	245853	83.3	15	2	AC125965 Rattus no
433		AC123270	247200	83.3	15	2	AC123270 Rattus no
434		AC114844	247294	83.3	15	2	AC114844 Rattus no
435	c	AL772318	249534	83.3	15	10	AL772318 Mouse DNA
436		AC095094	250774	83.3	15	2	AC095094 Rattus no
437		AC131820	251457	83.3	15	2	AC131820 Rattus no
438	c	AC127206	253241	83.3	15	2	AC127206 Rattus no
439		AC134361	254734	83.3	15	2	AC134361 Rattus no
440		AC095412	261901	83.3	15	2	AC095412 Rattus no
441	c	BX649488	270051	83.3	15	5	BX649488 Danio rer
442		AC127636	270715	83.3	15	2	AC127636 Rattus no
443		AC126593	280912	83.3	15	2	AC126593 Rattus no
444	c	AC106950	284590	83.3	15	2	AC106950 Rattus no
445	c	AC106283	289633	83.3	15	2	AC106283 Rattus no
446		AU028818	175	82.2	14.8	11	AU028818 Rattus no
447		AX439281	312	82.2	14.8	6	AX439281 Sequence
448		G71528	316	82.2	14.8	11	G71528 A6152453FM
449		BV073381	404	82.2	14.8	11	BV073381 S212P6646
450		AF117701	424	82.2	14.8	14	AF117701 Pestiviru
451		AF117702	424	82.2	14.8	14	AF117702 Pestiviru
452		AF302428	425	82.2	14.8	14	AF302428 Pestiviru
453		AF302429	425	82.2	14.8	14	AF302429 Pestiviru
454		AF302434	425	82.2	14.8	14	AF302434 Pestiviru
455		AF299317S2	439	82.2	14.8	14	AF299918 Pestiviru
456	c	AY358231	442	82.2	14.8	9	AY358231 Homo sapi
457		BD204306	453	82.2	14.8	6	BD204306 5'EST and
458		AX875525	559	82.2	14.8	6	AX875525 Sequence
459	c	BD155587	559	82.2	14.8	6	BD155587 Primer fo
460		G98495	581	82.2	14.8	11	G98495 S208P6397FE
461	c	AF160470	606	82.2	14.8	8	AF160470 Quisquali
462		AF160469	607	82.2	14.8	8	AF160469 Quisquali
463	c	AF425687	609	82.2	14.8	8	AF425687 Quisquali
464		CNS06FB2	625	82.2	14.8	11	AL396180 T7 end of
465	c	AF276131	641	82.2	14.8	5	AF276131 Scaphirhy
466	c	AF333702	653	82.2	14.8	5	AF333702 Ficus car
467		AY448152	658	82.2	14.8	9	AY448152 Cebus ape
468		AY448263	658	82.2	14.8	9	AY448263 Colobus g
469		AY448607	658	82.2	14.8	9	AY448607 Trachypit
470		AY449315	658	82.2	14.8	9	AY449315 Trachypit
471		AY454995	658	82.2	14.8	9	AY454995 Trachypit
472		AY455235	658	82.2	14.8	9	AY455235 Saimiri s
473	c	AF362401	702	82.2	14.8	14	AF362401 Peaton vi
474	c	BV025767	754	82.2	14.8	11	BV025767 S212P6042
475		GMU50150	771	82.2	14.8	8	U50150 Glycine max
476	c	AY048678	816	82.2	14.8	14	AY048678 Peaton vi
477		AY071329	829	82.2	14.8	3	AY071329 Drosophil
478		AY113164	942	82.2	14.8	8	AY113164 Arabidops
479		AF361631	1235	82.2	14.8	8	AF361631 Arabidops
480		AF145122	1371	82.2	14.8	8	AF145122 Homo sapi
481		AF116619	1393	82.2	14.8	17	AF116619 Homo sapi
482		AY150383	1405	82.2	14.8	8	AY150383 Arabidops
483		BC057669	1421	82.2	14.8	10	BC057669 Mus muscu
484		BC010828	1438	82.2	14.8	10	BC010828 Mus muscu
485		BC024876	1464	82.2	14.8	10	BC024876 Mus muscu
486		AF011925	1500	82.2	14.8	8	AF011925 Bos tauru
487		SCATP13	1529	82.2	14.8	8	X56215 Saccharomyc
488		AK113677	1670	82.2	14.8	3	AK113677 Ciona int
489		AX596870	1743	82.2	14.8	6	AX596870 Sequence
490		AX821296	1743	82.2	14.8	6	AX821296 Sequence
491*		AX832326	1743	82.2	14.8	6	AX832326 Sequence
492		E00968	1786	82.2	14.8	6	E00968 BanHI-HincI
493		AX078761	1878	82.2	14.8	6	AX078761 Sequence
494		BT004380	1909	82.2	14.8	8	BT004380 Arabidops
495		AY050856	1910	82.2	14.8	8	AY050856 Arabidops
496		HSH126112	2023	82.2	14.8	9	AL021682 Homo sapi
497		BT002944	2162	82.2	14.8	8	BT002944 Arabidops
498		AK055444	2197	82.2	14.8	9	AK055444 Homo sapi
499		AF239719	2254	82.2	14.8	8	AF239719 Arabidops
500		BC020660	2313	82.2	14.8	9	BC020660 Homo sapi

ALIGNMENTS

RESULT 1
```
LOCUS       BD274814                  18 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  CANCER CELL VACCINE.
ACCESSION   BD274814
VERSION     BD274814.1  GI:33084582
KEYWORDS    JP 2002531582-A/39.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
  TITLE     CANCER CELL VACCINE
  JOURNAL   Patent: JP 2002531582-A 39 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT     OS   Artificial Sequence
            PN   JP 2002531582-A/39
            PD   24-SEP-2002
            PF   24-NOV-1999 JP 2000058901
            PR   04-DEC-1998 US 09/205995
            PI   minzhen kusu,gang qiu,robert hunfrees
            CC   Description of Artificial Sequence: antisense oligonucleotide
            CC   corresponding
            CC   to a specific region of the mouse Ii gene.
            FH   Key             Location/Qualifiers
FEATURES             Location/Qualifiers
     source          1..18
                     /organism="unknown"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db          1 TTGGTCATCCATGGCTCT 18
```

RESULT 2
```
LOCUS       AR205280                  18 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 40 from patent US 6368855.
ACCESSION   AR205280
VERSION     AR205280.1  GI:21502824
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Xu,M., Qiu,G. and Humphreys,R.
  TITLE     MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
  JOURNAL   Patent: US 6368855-A 40 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..18
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db          1 TTGGTCATCCATGGCTCT 18
```

RESULT 3
```
LOCUS       I91647                    24 bp    DNA     linear   PAT 01-DEC-1998
DEFINITION  Sequence 1 from patent US 5726020.
ACCESSION   I91647
VERSION     I91647.1  GI:3936117
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Humphreys,R.E. and Xu,M.
  TITLE     Inhibition of II synthesis
  JOURNAL   Patent: US 5726020-A 1 10-MAR-1998;
FEATURES             Location/Qualifiers
     source          1..24
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db          3 TTGGTCATCCATGGCTCT 20
```

RESULT 4
```
LOCUS       BD274839/c               103 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  CANCER CELL VACCINE.
ACCESSION   BD274839
VERSION     BD274839.1  GI:33084607
KEYWORDS    JP 2002531582-A/64.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 103)
  AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
  TITLE     CANCER CELL VACCINE
  JOURNAL   Patent: JP 2002531582-A 64 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT     OS   Artificial Sequence
            PN   JP 2002531582-A/64
            PD   24-SEP-2002
            PF   24-NOV-1999 JP 2000058901
            PR   04-DEC-1998 US 09/205995
            PI   minzhen kusu,gang qiu,robert hunfrees
            CC   Description of Artificial Sequence: Reverse gene construct CC
            CC   corresponding to
            CC   a specific region of the mouse Ii gene.
            FH   Key             Location/Qualifiers
FEATURES             Location/Qualifiers
     source          1..103
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 103;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db         38 TTGGTCATCCATGGCTCT 21
```

RESULT 5
```
AR205305/c
LOCUS       AR205305
```

```
DEFINITION   Sequence 68 from patent US 6368855.
ACCESSION    AR205305
VERSION      AR205305.1  GI:21502855
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 103)
AUTHORS      Xu,M., Qiu,G. and Humphreys,R.
TITLE        MHC class II antigen presenting cells containing oligonucleotides
             which inhibit Ii protein expression
JOURNAL      Patent: US 6368855-A 68 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..103
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 103;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 TTGGTCATCCATGGCTCT 18
             ||||||||||||||||||
Db        38 TTGGTCATCCATGGCTCT 21
```

```
RESULT 6
BD274846/c
LOCUS        BD274846    107 bp    DNA    linear    PAT 17-JUL-2003
DEFINITION   CANCER CELL VACCINE.
ACCESSION    BD274846
VERSION      BD274846.1  GI:33084614
KEYWORDS     JP 2002531582-A/71.
SOURCE       synthetic construct
ORGANISM     synthetic construct
             artificial sequences.
REFERENCE    1  (bases 1 to 107)
AUTHORS      Kusu,M., Qiu,G. and Hunfrees,R.
TITLE        CANCER CELL VACCINE
JOURNAL      Patent: JP 2002531582-A 71 24-SEP-2002;
             ANTIGEN EXPRESS INC
COMMENT      OS   Artificial Sequence
             PN   JP 2002531582-A/71
             PD   24-SEP-2002
             PF   24-NOV-1999 JP 2000586901
             PR   04-DEC-1998 US 09/205995
             PI   minzhen kusu,gang qiu,robert hunfrees
             CC   Description of Artificial Sequence: Reverse gene construct CC
                  corresponding to
             CC   a specific region of the mouse Ii gene.
             FH   Key             Location/Qualifiers
     source          1..107
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 TTGGTCATCCATGGCTCT 18
             ||||||||||||||||||
Db       104 TTGGTCATCCATGGCTCT 87
```

```
RESULT 7
AR205312/c
LOCUS        AR205312    107 bp    DNA    linear    PAT 20-JUN-2002
DEFINITION   Sequence 75 from patent US 6368855.
ACCESSION    AR205312
VERSION      AR205312.1  GI:21502863
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 107)
AUTHORS      Xu,M., Qiu,G. and Humphreys,R.
TITLE        MHC class II antigen presenting cells containing oligonucleotides
             which inhibit Ii protein expression
JOURNAL      Patent: US 6368855-A 75 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..107
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 TTGGTCATCCATGGCTCT 18
             ||||||||||||||||||
Db       104 TTGGTCATCCATGGCTCT 87
```

```
RESULT 8
BD274850/c
LOCUS        BD274850    124 bp    DNA    linear    PAT 17-JUL-2003
DEFINITION   CANCER CELL VACCINE.
ACCESSION    BD274850
VERSION      BD274850.1  GI:33084618
KEYWORDS     JP 2002531582-A/75.
SOURCE       synthetic construct
ORGANISM     synthetic construct
             artificial sequences.
REFERENCE    1  (bases 1 to 124)
AUTHORS      Kusu,M., Qiu,G. and Hunfrees,R.
TITLE        CANCER CELL VACCINE
JOURNAL      Patent: JP 2002531582-A 75 24-SEP-2002;
             ANTIGEN EXPRESS INC
COMMENT      OS   Artificial Sequence
             PN   JP 2002531582-A/75
             PD   24-SEP-2002
             PF   24-NOV-1999 JP 2000586901
             PR   04-DEC-1998 US 09/205995
             PI   minzhen kusu,gang qiu,robert hunfrees
             CC   Description of Artificial Sequence: Reverse gene construct CC
                  corresponding to
             CC   a specific region of the mouse Ii gene.
             FH   Key             Location/Qualifiers
     source          1..124
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             100.0%;  Score 18;  DB 6;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 TTGGTCATCCATGGCTCT 18
             ||||||||||||||||||
Db        38 TTGGTCATCCATGGCTCT 21
```

```
RESULT 9
AR205316/c
LOCUS        AR205316    124 bp    DNA    linear    PAT 20-JUN-2002
DEFINITION   Sequence 79 from patent US 6368855.
ACCESSION    AR205316
VERSION      AR205316.1  GI:21502868
KEYWORDS     .
```

SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 124)
AUTHORS   Xu,M., Qiu,G. and Humphreys,R.
TITLE   MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
JOURNAL   Patent: US 6368855-A 79 09-APR-2002;
FEATURES   Location/Qualifiers
     source   1..124
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 18; DB 6; Length 124;
Best Local Similarity   100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db   38 TTGGTCATCCATGGCTCT 21

RESULT 10
I91650/c
LOCUS   I91650   141 bp   DNA   linear   PAT 01-DEC-1998
DEFINITION Sequence 4 from patent US 5726020.
ACCESSION   I91650
VERSION   I91650.1 GI:3936120
KEYWORDS   .
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 141)
AUTHORS   Humphreys,R.E. and Xu,M.
TITLE   Inhibition of Ii synthesis
JOURNAL   Patent: US 5726020-A 4 10-MAR-1998;
FEATURES   Location/Qualifiers
     source   1..141
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 18; DB 6; Length 141;
Best Local Similarity   100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db   59 TTGGTCATCCATGGCTCT 42

RESULT 11
BD274849/c
LOCUS   BD274849   148 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION CANCER CELL VACCINE.
ACCESSION   BD274849
VERSION   BD274849.1 GI:33084617
KEYWORDS   JP 2002531582-A/74.
SOURCE   synthetic construct
ORGANISM   synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 148)
AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
TITLE   CANCER CELL VACCINE
JOURNAL   Patent: JP 2002531582-A 74 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT   OS   Artificial Sequence
            PN   JP 2002531582-A/74
            PD   24-SEP-2002
            PF   24-NOV-1999 JP 2000586901
            PR   04-DEC-1998 US 09/205995
            PI   minzhen kusu,gang qiu,robert hunfrees

CC   Description of Artificial Sequence: Reverse gene construct CC
CC   corresponding to
CC   a specific region of the mouse Ii gene.
FH   Key   Location/Qualifiers
FEATURES
     source   1..148
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

Query Match   100.0%; Score 18; DB 6; Length 148;
Best Local Similarity   100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db   62 TTGGTCATCCATGGCTCT 45

RESULT 12
AR205315/c
LOCUS   AR205315   148 bp   DNA   linear   PAT 20-JUN-2002
DEFINITION Sequence 78 from patent US 6368855.
ACCESSION   AR205315
VERSION   AR205315.1 GI:21502867
KEYWORDS   .
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 148)
AUTHORS   Xu,M., Qiu,G. and Humphreys,R.
TITLE   MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
JOURNAL   Patent: US 6368855-A 78 09-APR-2002;
FEATURES   Location/Qualifiers
     source   1..148
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 18; DB 6; Length 148;
Best Local Similarity   100.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TTGGTCATCCATGGCTCT 18
              ||||||||||||||||||
Db   62 TTGGTCATCCATGGCTCT 45

RESULT 13
BD274843/c
LOCUS   BD274843   169 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION CANCER CELL VACCINE.
ACCESSION   BD274843
VERSION   BD274843.1 GI:33084611
KEYWORDS   JP 2002531582-A/68.
SOURCE   synthetic construct
ORGANISM   synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 169)
AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
TITLE   CANCER CELL VACCINE
JOURNAL   Patent: JP 2002531582-A 68 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT   OS   Artificial Sequence
            PN   JP 2002531582-A/68
            PD   24-SEP-2002
            PF   24-NOV-1999 JP 2000586901
            PR   04-DEC-1998 US 09/205995
            PI   minzhen kusu,gang qiu,robert hunfrees
            CC   Description of Artificial Sequence: Reverse gene construct CC
            CC   corresponding to

```
FEATURES             Location/Qualifiers
     source          1..169
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"

ORIGIN
Query Match             100.0%;  Score 18; DB 6;  Length 169;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db    104 TTGGTCATCCATGGCTCT 87


RESULT 14
AR205309/c
LOCUS       AR205309                 169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 72 from patent US 6368855.
ACCESSION   AR205309
VERSION     AR205309.1  GI:21502860
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 169)
AUTHORS     Xu,M., Qiu,G. and Humphreys,R.
TITLE       MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
JOURNAL     Patent: US 6368855-A 72 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..169
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN
Query Match             100.0%;  Score 18; DB 6;  Length 169;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db    104 TTGGTCATCCATGGCTCT 87


RESULT 15
BD274848/c
LOCUS       BD274848                 190 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  CANCER CELL VACCINE.
ACCESSION   BD274848
VERSION     BD274848.1  GI:33084616
KEYWORDS    JP 2002531582-A/73.
SOURCE      synthetic construct
ORGANISM    synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 190)
AUTHORS     Kusu,M., Qiu,G. and Hunfrees,R.
TITLE       CANCER CELL VACCINE
JOURNAL     Patent: JP 2002531582-A 73 24-SEP-2002;
            ANTIGEN EXPRESS INC
OS   Artificial Sequence
PN   JP 2002531582-A/73
PD   24-SEP-2002
PF   24-NOV-1999 JP 2000586901
PR   24-NOV-1999 JP 2000586901
PR   04-DEC-1998 US 09/205995
PI   minzhen kusu,gang qiu,robert hunfrees
CC   minzhen kusu.gang qiu.robert hunfrees
CC   Description of Artificial Sequence: Reverse gene construct CC
     corresponding to
CC   a specific region of the mouse Ii gene.
FH   Key             Location/Qualifiers


FEATURES             Location/Qualifiers
     source          1..190
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"

ORIGIN
Query Match             100.0%;  Score 18; DB 6;  Length 190;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db    104 TTGGTCATCCATGGCTCT 87


RESULT 16
AR205314/c
LOCUS       AR205314                 190 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 77 from patent US 6368855.
ACCESSION   AR205314
VERSION     AR205314.1  GI:21502866
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 190)
AUTHORS     Xu,M., Qiu,G. and Humphreys,R.
TITLE       MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
JOURNAL     Patent: US 6368855-A 77 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..190
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN
Query Match             100.0%;  Score 18; DB 6;  Length 190;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db    104 TTGGTCATCCATGGCTCT 87


RESULT 17
MMIIR/c
LOCUS       MMIIR                    252 bp    mRNA    linear   ROD 11-NOV-1988
DEFINITION  Mouse mIi mRNA for Ia-associated invariant chain.
ACCESSION   X07129
VERSION     X07129.1  GI:52637
KEYWORDS    antigen; Ia antigen-associated invariant chain.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Stone,J., Perry,R., Todd,J.A. and McDevitt,H.O.
TITLE       Nucleotide sequences of the murine Ia-Associated invariant chain
            (Ii) and I-E (H-2S, Beta) chain expressible cDNA clones
JOURNAL     Nucleic Acids Res. (1988) In press
REFERENCE   2  (bases 1 to 252)
AUTHORS     Todd,J.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-1988) Todd J., Department of Medical
            Microbiology, Stanford University, Stanford, CA 94305, USA
COMMENT     see x05428 - x05430 for overlapping sequence(s).
FEATURES             Location/Qualifiers
     source          1..252
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B10.S"
```

```
     CDS             79..>252
                     /db_xref="taxon:10090"
                     /clone="pmIiJS1"
                     /haplotype="H-2s"
                     /tissue_type="spleen"
                     /clone_lib="lambda gt10-JT85"
                     /codon_start=1
                     /note="unnamed protein product; mIi protein"
                     /protein_id="CAA30141.1"
                     /db_xref="GI:52638"
                     /db_xref="GOA:P04441"
                     /db_xref="SWISS-PROT:P04441"
                     /translation="MDDQRDLISNHEQLPILGNRPREPERCSRGALYTGVSVLVALLL
                     AGQATTAYFLYQQQ"
ORIGIN

Query Match   100.0%;  Score 18;  DB 10;  Length 252;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1  TTGGTCATCCATGGCTCT  18
          ||||||||||||||||||
Db    90  TTGGTCATCCATGGCTCT  73

RESULT 18
BV096208/c
LOCUS       BV096208       487 bp    DNA    linear    STS 15-OCT-2003
DEFINITION  RPAMMSEQ0008186 Roche Palo Alto Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV096208
VERSION     BV096208.1  GI:37673687
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
AUTHORS     Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
            McPherson,J.D., Foernzler,D. and Peltz,G.
TITLE       Mus musculus SNPs
JOURNAL     Unpublished (2003)
COMMENT     Contact: Jonathan Usuka
            Roche Palo Alto Genetics and Genomics Department
            Roche Palo Alto
            3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
            Tel: 6508555807
            Email: Jonathan.Usuka@roche.com
            Primer A: No primer submitted
            Primer B: No primer submitted.
FEATURES    Location/Qualifiers
     source          1..487
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /map="18-10042-10567-CAAA01071191.1.1.17884"
                     /clone_lib="Roche Palo Alto"
                     /note="SNPs developed from assay sequences derived from 15
                     different strains of mice (as of October 1, 2003). Those
                     strains include A/J, A/HeJ,~129/Sv, AKR/J, B10.D2-H2/oSnJ,
                     BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J,~CAST/Ei, DBA/2J,
                     MRL/Mp, NZB/BinJ, NZW/LaC, SPRET/Ei.~"
     STS             <1..>487
ORIGIN

Query Match   100.0%;  Score 18;  DB 11;  Length 487;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1  TTGGTCATCCATGGCTCT  18
          ||||||||||||||||||
Db    378  TTGGTCATCCATGGCTCT  361

RESULT 19
MUSIFNRES/c
LOCUS       MUSIFNRES       508 bp    DNA    linear    ROD 27-APR-1993
DEFINITION  Mouse MHC class II-associated invariant chain gene, exon 1.
ACCESSION   M35872
VERSION     M35872.1  GI:194129
KEYWORDS    MHC class II-associated invariant chain.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 508)
AUTHORS     Eades,A.M., Litfin,M. and Rahmsdorf,H.J.
TITLE       The IFN-gamma response of the murine invariant chain gene is
            mediated by a complex enhancer that includes several MHC class II
            consensus elements
JOURNAL     J. Immunol. 144 (11), 4399-4409 (1990)
MEDLINE     90257363
PUBMED      2111346
COMMENT     Original source text: Mouse DNA.
FEATURES    Location/Qualifiers
     source          1..508
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     misc_feature    87..158
                     /note="IFN-gamma response element"
     CAAT_signal     255..258
     misc_signal     267..272
                     /note="SP1 binding site"
     TATA_signal     315..319
     mRNA            348..>508
                     /product="invariant chain protein mRNA"
     CDS             432..>508
                     /note="invariant chain protein"
                     /number=1
                     /codon_start=1
                     /protein_id="AAA37897.1"
                     /db_xref="GI:553934"
                     /translation="MDDQRDLISNHEQLPILGNRPREPE"
ORIGIN

Query Match   100.0%;  Score 18;  DB 10;  Length 508;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1  TTGGTCATCCATGGCTCT  18
          ||||||||||||||||||
Db    443  TTGGTCATCCATGGCTCT  426

RESULT 20
MMIIGA/c
LOCUS       MMIIGA       585 bp    DNA    linear    ROD 24-FEB-1999
DEFINITION  Mouse Ii gene for Ia antigen associated invariant chains Ii31/Ii41
            and Ii41) exon 1.
ACCESSION   X05428
VERSION     X05428.1  GI:52626
KEYWORDS    alternative splicing; antigen; class II antigen; glycoprotein; Ia
            antigen; Ia antigen-associated invariant chain; repetitive
            sequence.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 585)
AUTHORS     Koch,N., Lauer,W., Habicht,J. and Dobberstein,B.
TITLE       Primary structure of the gene for the murine Ia antigen-associated
            invariant chains (Ii). An alternatively spliced exon encodes a
            cysteine-rich domain highly homologous to a repetitive sequence of
            thyroglobulin
```

```
JOURNAL     EMBO J. 6 (6), 1677-1683 (1987)
MEDLINE     87275861
PUBMED      3038530
COMMENT     see x05429-30 for remaining exons.
FEATURES             Location/Qualifiers
     source          1..585
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="AKR"
                     /db_xref="taxon:10090"
     misc_feature    22..36
                     /note="conserved class II MHC antigen 15-mer sequence"
     promoter        166..170
                     /note="pot. CAAT-box"
     misc_feature    178..183
                     /note="SpI protein binding site"
     promoter        226..230
                     /note="pot. TATA-box"
     misc_feature    230
                     /note="put. treanscription start site"
     repeat_region   237..312
                     /note="repetitive CAG sequence"
     mRNA            250..>419
     exon            250..>419
                     /number=1
     CDS             343..>419
                     /note="unnamed protein product; Ii protein"
                     /codon_start=1
                     /protein_id="CAA29010.1"
                     /db_xref="GI:52627"
                     /db_xref="GOA:P04441"
                     /db_xref="SWISS-PROT:P04441"
                     /translation="MDDQRDLISNHEQLPLIGNRPREPE"
     intron          420..>585
                     /number=1
ORIGIN

Query Match              100.0%;  Score 18;  DB 10;  Length 585;
Best Local Similarity    100.0%;  Pred. No. 31;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 TTGGTCATCCATGGCTCT 18
        ||||||||||||||||||
Db  354 TTGGTCATCCATGGCTCT 337

RESULT 21
BC003476/c
LOCUS       BC003476              1375 bp    mRNA    linear   ROD 03-OCT-2003
DEFINITION  Mus musculus Ia-associated invariant chain, mRNA (cDNA clone
            MGC:6517 IMAGE:2650401), complete cds.
ACCESSION   BC003476
VERSION     BC003476.1  GI:13097485
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1375)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE  22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1375)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: cgapbs-r@mail.nih.gov
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 5 Row: h Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1..1375
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:6517 IMAGE:2650401"
                     /tissue_type="Mammary tumor. Brca1-/fl; MMTV-Cre model. 10
                     months old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam3"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1375
                     /gene="Ii"
                     /note="synonyms: CD74, CLIP, DHLAG, HLADG, Ia-GAMMA"
                     /db_xref="LocusID:16149"
                     /db_xref="MGI:96534"
     CDS             86..925
                     /codon_start=1
                     /product="Ii protein"
                     /protein_id="AAH03476.1"
                     /db_xref="GI:13097486"
                     /db_xref="LocusID:16149"
                     /translation="MDDQRDLISNHEQLPILGNRPREPERCSRGALYTGVSLVIVALL
                     AGQATTAYFLYQQGRLDKLTITSQNLQLESLRMKLPKSAKPVSQMRMATPLLMRPMS
                     MDNMLLGPVKNVTKYGNMTQDHVMHLLTRSGPLEYPQLKGTPPENLKHLKNSMDGVNW
                     KIFESWMKQWLLFEMSKNSLEKKPTEAPPKVLTKCQEEVSHIPAVYPGAFRPKCDEN
                     GNYLPLQCHGSTGYCWCVFPNGTEVPHTKSRGRHNCSEPLDMEDLSSGLGVTRQELGQ
                     VTL"
     misc_feature    671..847
                     /note="thyroglobulin 1; Region: Thyroglobulin type-1
                     repeat. Thyroglobulin type 1 repeats are thought to be
                     involved in the control of proteolytic degradation. The
                     domain usually contains six conserved cysteines. These
                     form three disulphide bridges. Cysteines 1 pairs with 2, 3
                     with 4 and 5 with 6"
```

```
ORIGIN              /db_xref="CDD:pfam00086"

Query Match
Best Local Similarity    100.0%;    Score 18;  DB 10;  Length 1375;
Matches    18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 TTGGTCATCCATGGCTCT 18
        ||||||||||||||||||
Db   97 TTGGTCATCCATGGCTCT 80
```

RESULT 22
BC061489/c
```
LOCUS       BC061489              1576 bp    mRNA    linear   ROD 05-NOV-2003
DEFINITION  Mus musculus Ia-associated invariant chain, mRNA (cDNA clone
            MGC:70236 IMAGE:4922119), complete cds.
ACCESSION   BC061489
VERSION     BC061489.1  GI:38181505
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1576)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scherer,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Malek,J.A., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grinwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE       Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE     22388257
PUBMED      12477932
REFERENCE   2 (bases 1 to 1576)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 131 Row:i Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1576
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:70236 IMAGE:4922119"
                     /tissue_type="Salivary gland, 10 week old female mouse"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1576
                     /gene="Ii"
                     /note="synonyms: CD74, CLIP, DHLAG, HLADG, Ia-GAMMA"
                     /db_xref="LocusID:16149"
                     /db_xref="MGI:96534"
     CDS             425..1072
                     /codon_start=1
                     /product="Ia-associated invariant chain"
                     /protein_id="AAH61489.1"
                     /db_xref="GI:38181506"
                     /db_xref="LocusID:16149"
                     /translation="MDDQRDLISNHEQLPILGNRPREPERCSRGALYTGVSVLVALLI
                     AGQATTRAYFLYQQGGRLDKLITITSQNLQLESLRMKLPKSAKPVSQMRMATPLLMRPMS
                     MDNMLLGPVKNVTKYGNMTQDHVMHLLTRSGPLEYPQLKGTFPENLKHLKNSMDGVNW
                     KIFESWMKQWLLFEMSKNSLEKKPTEAPPKEPLDMEDLSSGLGVTRQELGQVTI"
ORIGIN

Query Match
Best Local Similarity    100.0%;    Score 18;  DB 10;  Length 1576;
Matches    18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 TTGGTCATCCATGGCTCT 18
        ||||||||||||||||||
Db  436 TTGGTCATCCATGGCTCT 419
```

RESULT 23
AR199681/c
```
LOCUS       AR199681              2580 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 23 from patent US 6355479.
ACCESSION   AR199681
VERSION     AR199681.1  GI:20249755
KEYWORDS    Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 2580)
AUTHORS     Webb,S.R., Wingvist,O., Karlsson,L., Jackson,M.R. and Peterson,P.A.
TITLE       MHC class II antigen-presenting systems and methods for activating
            CD4+ T cells
JOURNAL     Patent: US 6355479-A 23 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..2580
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Query Match
Best Local Similarity    100.0%;    Score 18;  DB 6;  Length 2580;
Matches    18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 TTGGTCATCCATGGCTCT 18
        ||||||||||||||||||
Db  463 TTGGTCATCCATGGCTCT 446
```

RESULT 24
MMIIG/c

```
LOCUS       MMIIG                   9658 bp    DNA     linear   ROD 29-OCT-1999
DEFINITION  Murine I gene for MHC class II(Ia) associated invariant chain.
ACCESSION   X13414
VERSION     X13414.1  GI:52625
KEYWORDS    I gene; invariant chain; major histocompatibility complex.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 9658)
AUTHORS     Zhu,L. and Jones,P.P.
TITLE       Complete sequence of the murine invariant chain (Ii) gene
JOURNAL     Nucleic Acids Res. 17 (1), 447-448 (1989)
MEDLINE     89098406
PUBMED      2492095
REFERENCE   2  (bases 1 to 9658)
AUTHORS     Jones,P.P.
TITLE       Direct Submission
JOURNAL     Submitted (31-OCT-1988) Jones P.P., Dept of Biological Sciences,
            Stanford University, Stanford, CA 94305 5020, USA
COMMENT     haplotype=k; library=cosmid.
FEATURES             Location/Qualifiers
     source          1..9658
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="AKR"
                     /db_xref="taxon:10090"
                     /chromosome="18"
                     /tissue_type="liver"
     CAAT_signal     757..761
     TATA_signal     817..821
     gene            845..9649
                     /gene="Ii"
     prim_transcript 845..9649
                     /gene="Ii"
     polyA_signal    9607..9612
                     /gene="Ii"
ORIGIN

Query Match        100.0%;  Score 18;  DB 10;  Length 9658;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches 18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db  945 TTGGTCATCCATGGCTCT 928
```

RESULT 25
AC139759

```
LOCUS       AC139759             203664 bp    DNA     linear   HTG 03-JUL-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-385E10, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC139759
VERSION     AC139759.3  GI:32441373
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 203664)
AUTHORS     Wilson,R.K.
TITLE       The sequence of Mus musculus clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 203664)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (12-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 203664)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (03-JUL-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            On Jul 3, 2003 this sequence version replaced gi:28372760.
COMMENT     ------------------- Genome Center -------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ------------ Project Information ------------
            Center project name: M_BA0385E10
            ----------- Summary Statistics -----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 202218 bases at least Q40
            Consensus quality: 202635 bases at least Q30
            Consensus quality: 203101 bases at least Q20
            Insert size: 144000; agarose-fp
            Insert size: 209824; sum-of-contigs
            Quality coverage: 13.63 in Q20 bases; agarose-fp
            Quality coverage: 12.16 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1         1174: contig of 1174 bp in length
            *    1175         1274: gap of unknown length
            *    1275         2288: contig of 1014 bp in length
            *    2289         2388: gap of unknown length
            *    2389        64695: contig of 62307 bp in length
            *   64696        64795: gap of unknown length
            *   64796       146302: contig of 81507 bp in length
            *  146303       146402: gap of unknown length
            *  146403       203664: contig of 57262 bp in length.
FEATURES             Location/Qualifiers
     source          1..203664
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-385E10"
     misc_feature    1..1174
                     /note="assembly_name:Contig11"
     misc_feature    1275..2288
                     /note="assembly_name:Contig22"
     misc_feature    2389..64695
                     /note="assembly_name:Contig23"
     misc_feature    64796..146302
                     /note="assembly_name:Contig24"
     misc_feature    146403..203664
                     /note="assembly_name:Contig25"
ORIGIN

Query Match        100.0%;  Score 18;  DB 2;  Length 203664;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches 18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||||||
Db  56519 TTGGTCATCCATGGCTCT 56536
```

RESULT 26

```
BD274827
LOCUS       BD274827                18 bp    DNA    linear  PAT 17-JUL-2003
DEFINITION  CANCER CELL VACCINE.
ACCESSION   BD274827
VERSION     BD274827.1  GI:33084595
KEYWORDS    .
SOURCE      synthetic construct
ORGANISM    synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 18)
  AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
  TITLE     CANCER CELL VACCINE
  JOURNAL   Patent: JP 2002531582-A 52 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT     OS  Artificial Sequence
            PN  JP 2002531582-A/52
            PD  24-SEP-2002
            PF  24-NOV-1999 JP 2000586901
            PR  04-DEC-1998 US 09/205995
            PI  minzhen kusu,gang qiu,robert hunfrees
            CC  Description of Artificial Sequence: antisense oligonucleotide
            CC  corresponding
            CC  to a specific region of the mouse Ii gene.
            FH  Key   Location/Qualifiers
FEATURES
     source 1..18
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
ORIGIN

  Query Match          94.4%; Score 17; DB 6; Length 18;
  Best Local Similarity 100.0%; Pred. No. 1.3e+02;
  Matches  17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy      1 TTGGTCATCCATGGCTC 17
          |||||||||||||||||
Db      2 TTGGTCATCCATGGCTC 18

RESULT 27
BD274828
LOCUS       BD274828                18 bp    DNA    linear  PAT 17-JUL-2003
DEFINITION  CANCER CELL VACCINE.
ACCESSION   BD274828
VERSION     BD274828.1  GI:33084596
KEYWORDS    .
SOURCE      synthetic construct
ORGANISM    synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 18)
  AUTHORS   Kusu,M., Qiu,G. and Hunfrees,R.
  TITLE     CANCER CELL VACCINE
  JOURNAL   Patent: JP 2002531582-A 53 24-SEP-2002;
            ANTIGEN EXPRESS INC
COMMENT     OS  Artificial Sequence
            PN  JP 2002531582-A/53
            PD  24-SEP-2002
            PF  24-NOV-1999 JP 2000586901
            PR  04-DEC-1998 US 09/205995
            PI  minzhen kusu,gang qiu,robert hunfrees
            CC  Description of Artificial Sequence: antisense oligonucleotide
            CC  corresponding
            CC  to a specific region of the mouse Ii gene.
            FH  Key   Location/Qualifiers
FEATURES
     source 1..18
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
ORIGIN

  Query Match          94.4%; Score 17; DB 6; Length 18;
```

```
  Best Local Similarity 100.0%; Pred. No. 1.3e+02;
  Matches  17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy      2 TGGTCATCCATGGCTCT 18
          |||||||||||||||||
Db      1 TGGTCATCCATGGCTCT 17

RESULT 28
AR205293
LOCUS       AR205293                18 bp    DNA    linear  PAT 20-JUN-2002
DEFINITION  Sequence 53 from patent US 6368855.
ACCESSION   AR205293
VERSION     AR205293.1  GI:21502840
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 18)
  AUTHORS   Xu,M., Qiu,G. and Humphreys,R.
  TITLE     MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
  JOURNAL   Patent: US 6368855-A 53 09-APR-2002;
FEATURES    Location/Qualifiers
     source 1..18
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

  Query Match          94.4%; Score 17; DB 6; Length 18;
  Best Local Similarity 100.0%; Pred. No. 1.3e+02;
  Matches  17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy      1 TTGGTCATCCATGGCTC 17
          |||||||||||||||||
Db      2 TTGGTCATCCATGGCTC 18

RESULT 29
AR205294
LOCUS       AR205294                18 bp    DNA    linear  PAT 20-JUN-2002
DEFINITION  Sequence 54 from patent US 6368855.
ACCESSION   AR205294
VERSION     AR205294.1  GI:21502841
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 18)
  AUTHORS   Xu,M., Qiu,G. and Humphreys,R.
  TITLE     MHC class II antigen presenting cells containing oligonucleotides
            which inhibit Ii protein expression
  JOURNAL   Patent: US 6368855-A 54 09-APR-2002;
FEATURES    Location/Qualifiers
     source 1..18
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

  Query Match          94.4%; Score 17; DB 6; Length 18;
  Best Local Similarity 100.0%; Pred. No. 1.3e+02;
  Matches  17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy      2 TGGTCATCCATGGCTCT 18
          |||||||||||||||||
Db      1 TGGTCATCCATGGCTCT 17

RESULT 30
BX890568/c
LOCUS       BX890568            146937 bp    DNA    linear  HTG 11-DEC-2003
DEFINITION  Danio rerio clone DKEY-20812, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
```

```
ACCESSION   BX890568
VERSION     BX890568.1  GI:39752493
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 146937)
AUTHORS     Button,J.
TITLE       Direct Submission
JOURNAL     Submitted (10-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     ------------ Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------ Project Information
            Center project name: zK20BJ2
            ------------ Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 138586 bases at least Q40
            Consensus quality: 140422 bases at least Q30
            Consensus quality: 141614 bases at least Q20
            Insert size: 144537; sum-of-contigs
            Insert size: 187037; 1.9% error; agarose-fp
            Quality coverage: 3.21x in Q20 bases; sum-of-contigs Quality
            coverage: 2.99x in Q20 bases; agarose-fp
            ------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1    2005: contig of 2005 bp in length
            *     2006    2105: gap of 100 bp
            *     2106    4727: contig of 2622 bp in length
            *     4728    4827: gap of 100 bp
            *     4828   12323: contig of 7496 bp in length
            *    12324   12423: gap of 100 bp
            *    12424   15294: contig of 2871 bp in length
            *    15295   15394: gap of 100 bp
            *    15395   18604: contig of 3210 bp in length
            *    18605   18704: gap of 100 bp
            *    18705   31814: contig of 13110 bp in length
            *    31815   31914: gap of 100 bp
            *    31915   36726: contig of 4812 bp in length
            *    36727   36826: gap of 100 bp
            *    36827   41989: contig of 5163 bp in length
            *    41990   42089: gap of 100 bp
            *    42090   45547: contig of 3458 bp in length
            *    45548   45647: gap of 100 bp
            *    45648   48163: contig of 2516 bp in length
            *    48164   48263: gap of 100 bp
            *    48264   50748: contig of 2485 bp in length
            *    50749   50848: gap of 100 bp
            *    50849   53746: contig of 2898 bp in length
            *    53747   53846: gap of 100 bp
            *    53847   56034: contig of 2188 bp in length
            *    56035   56134: gap of 100 bp
            *    56135   58624: contig of 2490 bp in length
            *    58625   58724: gap of 100 bp
            *    58725   62727: contig of 4003 bp in length
            *    62728   62827: gap of 100 bp
            *    62828   65027: contig of 2200 bp in length
            *    65028   65127: gap of 100 bp
            *    65128   68268: contig of 3141 bp in length
            *    68269   68368: gap of 100 bp
            *    68369   78970: contig of 10602 bp in length
            *    78971   79070: gap of 100 bp
            *    79071   83425: contig of 4355 bp in length
            *    83426   83525: gap of 100 bp
            *    83526   89946: contig of 6421 bp in length
            *    89947   90046: gap of 100 bp
            *    90047   95473: contig of 5427 bp in length
            *    95474   95573: gap of 100 bp
            *    95574  102377: contig of 6804 bp in length
            *   102378  102477: gap of 100 bp
            *   102478  115256: contig of 12779 bp in length
            *   115257  115356: gap of 100 bp
            *   115357  130709: contig of 15353 bp in length
            *   130710  130809: gap of 100 bp
            *   130810  146937: contig of 16128 bp in length
FEATURES             Location/Qualifiers
     source          1..146937
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-20BJ2"
                     /clone_lib="DanioKey"
     misc_feature    1..2005
                     /note="assembly_fragment:00150
                     fragment_chain:1"
     misc_feature    2106..4727
                     /note="assembly_fragment:00183
                     fragment_chain:1"
     misc_feature    4828..12323
                     /note="assembly_fragment:00464
                     fragment_chain:1"
     misc_feature    12424..15294
                     /note="assembly_fragment:00272
                     fragment_chain:1"
     misc_feature    15395..18604
                     /note="assembly_fragment:00342
                     fragment_chain:1"
     misc_feature    18705..31814
                     /note="assembly_fragment:00724
                     fragment_chain:2"
     misc_feature    31915..36726
                     /note="assembly_fragment:00322
                     fragment_chain:2"
     misc_feature    36827..41989
                     /note="assembly_fragment:00537
                     fragment_chain:2"
     misc_feature    42090..45547
                     /note="assembly_fragment:00258
                     fragment_chain:2"
     misc_feature    45648..48163
                     /note="assembly_fragment:00219
                     fragment_chain:3"
     misc_feature    48264..50748
                     /note="assembly_fragment:00195
                     fragment_chain:3"
     misc_feature    50849..53746
                     /note="assembly_fragment:00288
                     fragment_chain:3"
     misc_feature    53847..56034
                     /note="assembly_fragment:00132
                     fragment_chain:4"
     misc_feature    56135..58624
                     /note="assembly_fragment:00161
                     fragment_chain:4"
     misc_feature    58725..62727
                     /note="assembly_fragment:00232
                     fragment_chain:5"
     misc_feature    62828..65027
                     /note="assembly_fragment:00141
                     fragment_chain:5"
     misc_feature    65128..68268
                     /note="assembly_fragment:00304
```

```
misc_feature        fragment_chain:6"
                    68369..78970
                    /note="assembly_fragment:00576"
                    fragment_chain:6"
misc_feature        79071..83425
                    /note="assembly_fragment:00370"
misc_feature        83326..83946
                    /note="assembly_fragment:00400"
misc_feature        90047..95473
                    /note="assembly_fragment:00431"
misc_feature        95574..102377
                    /note="assembly_fragment:00498.0"
misc_feature        102478..115256
                    /note="assembly_fragment:00644"
misc_feature        115357..130709
                    /note="assembly_fragment:00816"
misc_feature        130810..146937
                    /note="assembly_fragment:00925.0"

ORIGIN

Query Match         94.4%;   Score 17;  DB 2; Length 146937;
Best Local Similarity  100.0%;  Pred. No. 1.2e+02;
Matches   17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      2 TGGTCATCCATGGCTCT 18
          |||||||||||||||||
Db  13202 TGGTCATCCATGGCTCT 13186


RESULT 31
AC121176/c
LOCUS       AC121176           183607 bp    DNA    linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-435M17, WORKING DRAFT SEQUENCE.
ACCESSION   AC121176
VERSION     AC121176.5  GI:25137867
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 183607)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
            Allen,C., Allen.H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi.V., Aoyagi,A., Ayodeji.M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,J., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Mitja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regler,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,R., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 183607)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 183607)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23908843.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------- Project Information
                Center project name: GWDB
                Center clone name: CH230-435M17
            ------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 174081 bases at least Q40
                Consensus quality: 174876 bases at least Q30
                Consensus quality: 175367 bases at least Q20
                Estimated insert size: 178972; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ------------------------------------------------------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. Gaps between the contigs
            *   are represented as runs of N. The order of the pieces
            *   is believed to be correct as given, however the sizes
            *   of the gaps between them are based on estimates that have
            *   provided by the submittor.
            * This sequence will be replaced
            *   by the finished sequence as soon as it is available and
            *   the accession number will be preserved.
            *       1  183607: contig of 183607 bp in length.
```

```
FEATURES     Location/Qualifiers
source       1..183607
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10116"
             /clone="CH230-435M17"
misc_feature 1..1466
             /note="wgs end extension
             clone_end:T7"
misc_feature 2430..3370
             /note="clone_boundary
             clone_end:T7
             site:
             end_sequence:BZ185307"
misc_feature complement(172775..173610)
             /note="clone_boundary
             clone_end:Sp6
             site:
             end_sequence:BZ185308"

ORIGIN

Query Match           94.4%;    Score 17;  DB 2;  Length 183607;
Best Local Similarity 100.0%;   Pred. No. 1.2e+02;
Matches   17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 TTGGTCATCCATGGCTC 17
           ||||||||||||||||||
Db  150176 TTGGTCATCCATGGCTC 150160


RESULT 32
AC126656
LOCUS       AC126656   231571 bp  DNA  linear  HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-98L14, WORKING DRAFT SEQUENCE.
ACCESSION   AC126656
VERSION     AC126656.4  GI:30581493
KEYWORDS    HTG; HTGS_PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231571)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,R.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Forbes,L., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,R., Grady,M., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,K., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,X., Zhou,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 231571)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (08-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231571)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23664704.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------- Project Information
                Center project name: GKOU
                Center clone name: CH230-98L14
            ------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 222856 bases at least Q40
                Consensus quality: 224263 bases at least Q30
                Consensus quality: 225265 bases at least Q20
                Estimated insert size: 231564; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. Gaps between the contigs
            *   are represented as runs of N. The order of the pieces
            *   is believed to be correct as given, however the sizes
            *   of the gaps between them are based on estimates that have
            *   provided by the submittor.
            * This sequence will be replaced
            *   by the finished sequence as soon as it is available and
            *   the accession number will be preserved.
```

```
FEATURES
    *               1    231571: contig of 231571 bp in length.
    source          Location/Qualifiers
                    1..231571
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-98L14"
    misc_feature    1..1407
                    /note="wgs_end_extension
                    clone_end:T7"
    misc_feature    complement(8623..9421)
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH310512"

ORIGIN

Query Match        94.4%;   Score 17;   DB 2;   Length 231571;
Best Local Similarity   100.0%;   Pred. No. 1.2e+02;
Matches    17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy          2  TGGTCATCCATGGCTCT 18
               |||||||||||||||||
Db     130618  TGGTCATCCATGGCTCT 130634


RESULT 33
AC111384/c
LOCUS       AC111384          239467 bp    DNA      linear    HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-138J5, WORKING DRAFT SEQUENCE, 7
            unordered pieces.
ACCESSION   AC111384
VERSION     AC111384.5  GI:30578946
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 239467)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,N., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,L., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagcon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,-L.-L.,
            Puazo,M., Quiroz,J., Racalin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,M., Rose,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 239467)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 239467)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24942349.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GMDP
            Center clone name: CH230-138J5
            ----- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 218505 bases at least Q40
            Consensus quality: 220520 bases at least Q30
            Consensus quality: 222019 bases at least Q20
            Estimated insert size: 222300; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *       consists of 7 contigs. The true order of the pieces
          *       is not known and their order in this sequence record is
          *       arbitrary. Gaps between the contigs are represented as
          *       runs of N, but the exact sizes of the gaps are unknown.
          *       This record will be updated with the finished sequence
          *       as soon as it is available and the accession number will
          *       be preserved.
          *       1        3374: contig of 3374 bp in length
          *       3375     3474: gap of unknown length
          *       3475   233285: contig of 229811 bp in length
          *     233286   233385: gap of unknown length
```

```
*      233386     234467: contig of 1082 bp in length
*      234468     234567: gap of unknown length
*      234568     235719: contig of 1152 bp in length
*      235720     235819: gap of unknown length
*      235820     237255: contig of 1436 bp in length
*      237256     237355: gap of unknown length
*      237356     238364: contig of 1009 bp in length
*      238365     238464: gap of unknown length
*      238465     239467: contig of 1003 bp in length.
FEATURES           Location/Qualifiers
     source        1..239467
                   /organism="Rattus norvegicus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:10116"
                   /clone="CH230-13B05"
     misc_feature  1..1245
                   /note="wgs end_extension
                   clone_end:T7"
     misc_feature  2890..3669
                   /note="clone_boundary
                   clone_end:T7
                   site:EcoRI
                   end_sequence:BZ112270"
     misc_feature  228381..230369
                   /note="wgs_contig"
     misc_feature  230420..233285
                   /note="wgs_contig"
ORIGIN

Query Match          94.4%;  Score 17;  DB 2;  Length 239467;
Best Local Similarity  100.0%;  Pred. No. 1.2e+02;
Matches  17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTC 17
          |||||||||||||||||
Db  80625 TTGGTCATCCATGGCTC 80609

RESULT 34
AX795517
LOCUS       AX795517               521 bp    DNA     linear   PAT 04-OCT-2003
DEFINITION  Sequence 86 from Patent WO03052141.
ACCESSION   AX795517
VERSION     AX795517.1  GI:37516184
KEYWORDS    .
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1
  AUTHORS   van Eijk,M.J. and Hogers,R.C.
  TITLE     High throughput analysis and detection of multiple target sequences
            using circular probes
  JOURNAL   Patent: WO 03052141-A 86 26-JUN-2003;
            Keygene N.V. (NL)
FEATURES           Location/Qualifiers
     source        1..521
                   /organism="Lycopersicon esculentum"
                   /mol_type="unassigned DNA"
                   /db_xref="taxon:4081"
ORIGIN

Query Match          91.1%;  Score 16.4;  DB 6;  Length 521;
Best Local Similarity  94.4%;  Pred. No. 2.8e+02;
Matches  17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||| |||
Db    357 TTGGTCATCCATGGTTCT 374

RESULT 35
```

```
AX796809
LOCUS       AX796809               521 bp    DNA     linear   PAT 04-OCT-2003
DEFINITION  Sequence 86 from Patent WO03052142.
ACCESSION   AX796809
VERSION     AX796809.1  GI:37517463
KEYWORDS    .
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1
  AUTHORS   van Eijk,M.J. and Hogers,R.C.
  TITLE     Analysis and detection of multiple target sequences using circular
            probes
  JOURNAL   Patent: WO 03052142-A 86 26-JUN-2003;
            Keygene N.V. (NL)
FEATURES           Location/Qualifiers
     source        1..521
                   /organism="Lycopersicon esculentum"
                   /mol_type="unassigned DNA"
                   /db_xref="taxon:4081"
ORIGIN

Query Match          91.1%;  Score 16.4;  DB 6;  Length 521;
Best Local Similarity  94.4%;  Pred. No. 2.8e+02;
Matches  17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||| |||
Db    357 TTGGTCATCCATGGTTCT 374

RESULT 36
AX797071
LOCUS       AX797071               521 bp    DNA     linear   PAT 04-OCT-2003
DEFINITION  Sequence 86 from Patent WO03052140.
ACCESSION   AX797071
VERSION     AX797071.1  GI:37517724
KEYWORDS    .
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1
  AUTHORS   van Eijk,M.J. and Hogers,R.C.
  TITLE     High throughput analysis and detection of multiple target sequences
  JOURNAL   Patent: WO 03052140-A 86 26-JUN-2003;
            Keygene N.V. (NL)
FEATURES           Location/Qualifiers
     source        1..521
                   /organism="Lycopersicon esculentum"
                   /mol_type="unassigned DNA"
                   /db_xref="taxon:4081"
ORIGIN

Query Match          91.1%;  Score 16.4;  DB 6;  Length 521;
Best Local Similarity  94.4%;  Pred. No. 2.8e+02;
Matches  17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          ||||||||||||||| |||
Db    357 TTGGTCATCCATGGTTCT 374

RESULT 37
AL603822
LOCUS       AL603822             42433 bp    DNA     linear   ROD 19-APR-2002
DEFINITION  Mouse DNA sequence from clone RP23-428P14 on chromosome 11,
            complete sequence.
ACCESSION   AL603822
VERSION     AL603822.9  GI:20268837
```

KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Harrison,E.
TITLE   Direct Submission
JOURNAL   Submitted (19-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Apr 22, 2002 this sequence version replaced gi:20196593.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-428P14 is
            from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES   Location/Qualifiers
     source   1..42433
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="RP23-428P14"
                     /clone_lib="RPCI-23"
ORIGIN

Query Match   91.1%;   Score 16.4;   DB 10;   Length 42433;
Best Local Similarity 94.4%;   Pred. No. 2.8e+02;
Matches 17; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 TTGGTCATCCATGGCTCT 18
             ||||||| ||||||||||
Db   20947 TTGGTCATCCGTGGCTCT 20964

RESULT 38
AC023361
LOCUS   AC023361   79271 bp   DNA   linear   HTG 16-OCT-2001
DEFINITION Mus musculus clone RP23-316K16, *** SEQUENCING IN PROGRESS ***, 21
            unordered pieces.
ACCESSION   AC023361
VERSION   AC023361.5 GI:16118050
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 79271)
AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 79271)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (14-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Oct 14, 2001 this sequence version replaced gi:11079344.
            -------------- Genome Center

            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information

            Center project name: MAAE
            Center clone name: RP23-316K16
            -------------- Summary Statistics

            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy: 74% of reads
            Chemistry: Dye-terminator Big Dye: 26% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 79638 bases at least Q40
            Consensus quality: 100663 bases at least Q30
            Consensus quality: 110213 bases at least Q20
            Estimated insert size: 101728; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 1.2x in Q20 bases; sum-of-contigs estimation
            ----
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 21 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1   8407   contig of 8407 bp in length
          *   8408   8507   gap of unknown length
          *   8508   14631   contig of 6124 bp in length
          *   14632   14731   gap of unknown length
          *   14732   18775   contig of 4044 bp in length
          *   18776   18875   gap of unknown length
          *   18876   23354   contig of 4479 bp in length
          *   23355   23454   gap of unknown length
          *   23455   29014   contig of 5560 bp in length
          *   29015   29114   gap of unknown length
          *   29115   33875   contig of 4761 bp in length
          *   33876   33975   gap of unknown length
          *   33976   39321   contig of 5346 bp in length
          *   39322   39421   gap of unknown length
          *   39422   42432   contig of 3011 bp in length
          *   42433   42532   gap of unknown length
          *   42533   45875   contig of 3343 bp in length
          *   45876   45975   gap of unknown length
          *   45976   49070   contig of 3095 bp in length
          *   49071   49170   gap of unknown length
          *   49171   52556   contig of 3386 bp in length
          *   52557   52656   gap of unknown length
          *   52657   55765   contig of 3109 bp in length
          *   55766   55865   gap of unknown length
          *   55866   59212   contig of 3347 bp in length
          *   59213   59312   gap of unknown length

```
*        59313   61392: contig of 2080 bp in length
*        61393   61492: gap of unknown length
*        61493   64355: contig of 2863 bp in length
*        64356   64455: gap of unknown length
*        64456   66618: contig of 2163 bp in length
*        66619   66718: gap of unknown length
*        66719   68868: contig of 2150 bp in length
*        68869   68968: gap of unknown length
*        68969   72179: contig of 3211 bp in length
*        72180   72279: gap of unknown length
*        72280   74374: contig of 2095 bp in length
*        74375   74474: gap of unknown length
*        74475   77052: contig of 2578 bp in length
*        77053   77152: gap of unknown length
*        77153   79271: contig of 2119 bp in length.
FEATURES          Location/Qualifiers
     source       1..79271
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /clone="RP23-316K16"
ORIGIN
     Query Match           91.1%;   Score 16.4;   DB 2;   Length 79271;
     Best Local Similarity 94.4%;   Pred. No. 2.8e+02;
     Matches 17;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     1 TTGGTCATCCATGGCTCT 18
         |||||||||| ||||||
Db 24001 TTGGTCATCCGTGGCTCT 24018

RESULT 39
AL928696/c
LOCUS       AL928696    102448 bp    DNA    linear   ROD 21-NOV-2002
DEFINITION  Mouse DNA sequence from clone RP23-387G11 on chromosome 2, complete
            sequence.
ACCESSION   AL928696
VERSION     AL928696.6  GI:25168716
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 102448)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 22, 2002 this sequence version replaced gi:25136705.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence. Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
```

```
*       assembly was confirmed by restriction digest, except on the rare
*       occasion of the clone being a YAC.
        The following abbreviations are used to associate primary accession
        numbers given in the feature table with their source databases:
        Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
        on the WORMPEP database can be found at:
        http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-387G11 is
        from the RPCI-23 Mouse PAC library
        constructed by the group of Pieter de Jong.
        For further details see http://www.chori.org/bacpac/home.htm
        VECTOR: pBACe3.6.
FEATURES          Location/Qualifiers
     source       1..102448
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /chromosome="2"
                  /clone="RP23-387G11"
                  /clone_lib="RPCI-23"
ORIGIN
     Query Match           91.1%;   Score 16.4;   DB 10;   Length 102448;
     Best Local Similarity 94.4%;   Pred. No. 2.8e+02;
     Matches 17;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     1 TTGGTCATCCATGGCTCT 18
         |||||||||| ||||||
Db 47844 TTGGTCATCCCTGGCTCT 47827

RESULT 40
AC114395
LOCUS       AC114395    113950 bp    DNA    linear   HTG 27-MAR-2002
DEFINITION  Zea mays chromosome unknown clone ZM06E22, *** SEQUENCING IN
            PROGRESS ***, 9 ordered pieces.
ACCESSION   AC114395
VERSION     AC114395.2  GI:19745056
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 113950)
  AUTHORS   Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,
            Gansberger,K. and Buell,R.
  TITLE     Zea mays BAC clone ZM06E22 BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113950)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 113950)
  AUTHORS   Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,
            Gansberger,K. and Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) University of Wisconsin, Department of
            Horticulture, Madison, WI 53706, USA
COMMENT     On Mar 27, 2002 this sequence version replaced gi:19263254.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * been provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *      1   12871: contig of 12871 bp in length
            *  12872   12971: gap of unknown length
            *  12972   26528: contig of 13557 bp in length
            *  26529   26628: gap of unknown length
```

```
        *       26629    27837: contig of 1209 bp in length
        *       27838    27937: gap of unknown length
        *       27938    33890: contig of 5953 bp in length
        *       33891    33990: gap of unknown length
        *       33991    35411: contig of 1421 bp in length
        *       35412    35511: gap of unknown length
        *       35512    36079: contig of 568 bp in length
        *       36080    36179: gap of unknown length
        *       36180    77297: contig of 41118 bp in length
        *       77298    77397: gap of unknown length
        *       77398    86122: contig of 8725 bp in length
        *       86123    86222: gap of unknown length
        *       86223   113950: contig of 27728 bp in length.

FEATURES             Location/Qualifiers
     source          1..113950
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="unknown"
                     /clone="ZM06E22"
                     /note="The BAC contains DNA sequences associated with
                     maize centromeres."

ORIGIN

Query Match         91.1%;  Score 16.4;  DB 2;  Length 113950;
Best Local Similarity 94.4%;  Pred. No. 2.8e+02;
Matches   17; Conservative   0; Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          |||||||| |||||||||
Db  87505 TTGGTCATACATGGCTCT 87522

RESULT 41
AC114395/c
LOCUS       AC114395              113950 bp    DNA     linear   HTG 27-MAR-2002
DEFINITION  Zea mays chromosome unknown clone ZM06E22, *** SEQUENCING IN
            PROGRESS ***, 9 ordered pieces.
ACCESSION   AC114395
VERSION     AC114395.2  GI:19745056
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 113950)
AUTHORS     Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,
            Gansberger,K. and Buell,R.
TITLE       Zea mays BAC clone ZM06E22 BAC genomic sequence
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 113950)
AUTHORS     Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,
            Buell,R.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAR-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 113950)
AUTHORS     Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,
            Gansberger,K. and Buell,R.
TITLE       Direct Submission
JOURNAL     Submitted (27-MAR-2002) University of Wisconsin, Department of
            Horticulture, Madison, WI 53706, USA
COMMENT     On Mar 27, 2002 this sequence version replaced gi:19263254.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1  12871: contig of 12871 bp in length
```

```
        *       12872    12971: gap of unknown length
        *       12972    26528: contig of 13557 bp in length
        *       26529    26628: gap of unknown length
        *       26629    27837: contig of 1209 bp in length
        *       27838    27937: gap of unknown length
        *       27938    33890: contig of 5953 bp in length
        *       33891    33990: gap of unknown length
        *       33991    35411: contig of 1421 bp in length
        *       35412    35511: gap of unknown length
        *       35512    36079: contig of 568 bp in length
        *       36080    36179: gap of unknown length
        *       36180    77297: contig of 41118 bp in length
        *       77298    77397: gap of unknown length
        *       77398    86122: contig of 8725 bp in length
        *       86123    86222: gap of unknown length
        *       86223   113950: contig of 27728 bp in length.

FEATURES             Location/Qualifiers
     source          1..113950
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="unknown"
                     /clone="ZM06E22"
                     /note="The BAC contains DNA sequences associated with
                     maize centromeres."

ORIGIN

Query Match         91.1%;  Score 16.4;  DB 2;  Length 113950;
Best Local Similarity 94.4%;  Pred. No. 2.8e+02;
Matches   17; Conservative   0; Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          |||||||| |||||||||
Db  42710 TTGGTCATACATGGCTCT 42693

RESULT 42
AC132215/c
LOCUS       AC132215              126323 bp    DNA     linear   PLN 03-SEP-2002
DEFINITION  Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0076E06, from chromosome 3, complete sequence.
ACCESSION   AC132215
VERSION     AC132215.1  GI:22657512
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 126323)
AUTHORS     McCombie,W.R., de la Bastide,M., Spiegel,L., Preston,R.,
            Nascimento,L., Zutavern,T., Balija,V., Bell,M., Miller,B.,
            Katzenberger,F., Muller,S., Sullivan,P., Yang,C., Dike,S.,
            O'Shaughnessy,A., Palmer,L. and Dedhia,N.
TITLE       Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0076E06, from chromosome 3, complete sequence
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 126323)
AUTHORS     McCombie,W.R.
TITLE       Direct Submission
JOURNAL     Submitted (03-SEP-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The nucleotide
            sequence of this BAC clone was generated by combining Syngenta,
            Monsanto and Cold Spring Harbor Laboratory Genome Center sequencing
            data.
```

```
FEATURES             Location/Qualifiers
     source          1..126323
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OSJNBa0076E06"
                     /clone_lib="HindIII"
     misc_feature    4490..4506
                     /note="We believe the assembly to be correct. The
                     sequence is a mononucleotide (C) repeat in which the exact
                     number of Cs is unknown. The majority of subclones in
                     the area agree with the assembly, however, two subclones
                     show two additional Cs."
     misc_feature    91540..91640
                     /note="We believe the assembly to be correct. The
                     sequence is covered by a PCR product which was amplified
                     with a high fidelity polymerase. The sequence is of high
                     quality and there is partial coverage by several subclones
                     with quality below phred30."
ORIGIN

Query Match     91.1%;    Score 16.4;    DB 8;    Length 126323;
Best Local Similarity  94.4%;    Pred. No. 2.8e+02;
Matches  17;    Conservative  0;    Mismatches  1;    Indels  0;    Gaps  0;

Qy       1 TTGGTCATCCATGGCTCT 18
           |||||||||||||||||
Db   21558 TTGGTCATCCATTGCTCT 21541

RESULT 43
AC025192
LOCUS       AC025192              134514 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens chromosome 8 clone RP11-509E2 map 8, 3 unordered
            pieces.
ACCESSION   AC025192
VERSION     AC025192.4  GI:11875263
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134514)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134514)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Cooke,C., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,J., Liu,G., Locke,K., McGurk,A., McKernan,K., McPheeters,R.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., Oliver,J., Peterson,K., Pierre,N.,
            O'Neil,D., Olivar,T.M., Oliver,J., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Pisani,C., Pollara,V., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 134514)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Norman,C.H., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 25, 2002 this sequence version replaced gi:12313839.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------------ Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ----------------------- Project Information
                Center project name: L5592
                Center clone name: 509_E_2
            -----------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record as
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     17787: contig of 17787 bp in length
            *  17788     17887: gap of 100 bp
            *  17888     97188: contig of 79301 bp in length
            *  97189     97288: gap of 100 bp
            *  97289    134514: contig of 37226 bp in length
FEATURES             Location/Qualifiers
     source          1..134514
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-509E2"
                     /clone_lib="RPCI-11 Human Male BAC"
ORIGIN

Query Match     91.1%;    Score 16.4;    DB 2;    Length 134514;
Best Local Similarity  94.4%;    Pred. No. 2.8e+02;
Matches  17;    Conservative  0;    Mismatches  1;    Indels  0;    Gaps  0;

Qy       1 TTGGTCATCCATGGCTCT 18
           ||||||||||||||||||
Db   95549 TTGGTCATCCATGGCTCT 95566

RESULT 44
```

```
AF248716
LOCUS       AF248716    146306 bp    DNA    linear    HTG 12-APR-2000
DEFINITION  Mus musculus chromosome 11 clone CT7-327023, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AF248716
VERSION     AF248716.1  GI:7542829
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 146306)
AUTHORS     Loots,G.G., Locksley,R.M., Blankespoor,C.M., Wang,Z.E., Miller,W.,
            Rubin,E.M. and Frazer,K.A.
TITLE       Identification of a coordinate regulator of interleukins 4, 13, and
            5 by cross-species sequence comparisons
JOURNAL     Science 288 (5463), 136-140 (2000)
MEDLINE     20217223
PUBMED      10753117
REFERENCE   2  (bases 1 to 146306)
AUTHORS     Dean,W.B., Lewis,K.D., Blankespoor,C.M., Nyugen,A., Loots,G.G.,
            Rubin,E.M. and Frazer,K.A.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAR-2000) Genome Sciences Department, Lawrence
            Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA
            94720, USA
COMMENT     Note: This is a 'working draft' sequence. It consists of 8 contigs.
            The order of contigs is believed to be correct. Gaps between the
            contigs are represented by a string of 10 Ns and the exact sizes of
            the gaps are unknown.
                    1 -  2488:  contig of 2488 bp in length
                 2489 -  2498:  gap of unknown length
                 2499 - 49928:  contig of 47429 bp in length
                49929 - 49937:  gap of unknown length
                49938 - 80580:  contig of 30643 bp in length
                80581 - 80590:  gap of unknown length
                80590 - 100725: contig of 20135 bp in length
               100726 - 100735: gap of unknown length
               100736 - 117337: contig of 16602 bp in length
               117338 - 117347: gap of unknown length
               117348 - 122392: contig of 5054 bp in length
               122393 - 122402: gap of unknown length
               122403 - 131369: contig of 8967 bp in length
               131370 - 131379: gap of unknown length
               131380 - 146306: contig of 14927 bp in length.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES    Location/Qualifiers
source      1..146306
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="11"
            /clone="CT7-327023"
ORIGIN

Query Match        91.1%;   Score 16.4;   DB 2;   Length 146306;
Best Local Similarity   94.4%;   Pred. No. 2.8e+02;
Matches  17;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy      1 TTGGTCATCCATCGCTCT 18
          ||||||||| ||||||||
Db  79479 TTGGTCATCCGTGGCTCT 79496

RESULT 45
AC140494/c
LOCUS       AC140494    151820 bp    DNA    linear    HTG 23-FEB-2003
DEFINITION  Homo sapiens chromosome UNK clone RP13-1034D17, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC140494
```

```
VERSION     AC140494.1  GI:28475768
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151820)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 151820)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     ----------------------  Genome Center  ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ----------------------  Project Information  ----------------------
            Center project name: H_FH1034D17
            ----------------------  Summary Statistics  ----------------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 147527 bases at least Q40
            Consensus quality: 148411 bases at least Q30
            Consensus quality: 148866 bases at least Q20
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      2172: contig of 2172 bp in length
            *     2173      2272: gap of unknown length
            *     2273      4219: contig of 1947 bp in length
            *     4220      4319: gap of unknown length
            *     4320      7107: contig of 2788 bp in length
            *     7108      7207: gap of unknown length
            *     7208     15791: contig of 8584 bp in length
            *    15792     15891: gap of unknown length
            *    15892     30772: contig of 14881 bp in length
            *    30773     30872: gap of unknown length
            *    30873     50525: contig of 19653 bp in length
            *    50526     50625: gap of unknown length
            *    50626     81756: contig of 31131 bp in length
            *    81757     81856: gap of unknown length
            *    81857    104940: contig of 23084 bp in length
            *   104941    105040: gap of unknown length
            *   105041    151820: contig of 46780 bp in length.
FEATURES    Location/Qualifiers
source      1..151820
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="UNK"
            /clone="RP13-1034D17"
misc_feature   1..2172
               /note="assembly_name:Contig8"
misc_feature   2273..4219
               /note="assembly_name:Contig9"
misc_feature   4320..7107
               /note="assembly_name:Contig10"
```

```
     misc_feature    7208. .15791
                     /note="assembly_name:Contig11"
     misc_feature    15892. .30772
                     /note="assembly_name:Contig12"
     misc_feature    30873. .50525
                     /note="assembly_name:Contig13"
     misc_feature    50626. .81756
                     /note="assembly_name:Contig14"
     misc_feature    81857. .104940
                     /note="assembly_name:Contig15"
     misc_feature    105041. .151820
                     /note="assembly_name:Contig16"
ORIGIN

Query Match       91.1%;    Score 16.4;  DB 2;   Length 151820;
Best Local Similarity   94.4%;   Pred. No. 2.8e+02;
Matches   17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          |||| |||||||||||||
Db  41448 TTGGCCATCCATGGCTCT 41431

RESULT 46
AC138308
LOCUS       AC138308          153988 bp   DNA    linear   HTG 22-DEC-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-286G16, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC138308
VERSION     AC138308.1  GI:27356763
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 153988)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       The sequence of Mus musculus clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 153988)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (21-DEC-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 153988)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (22-DEC-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     ----------- Genome Center -----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ----------- Project Information -----------
            Center project name: M_BA0286G16
            ----------- Summary Statistics -----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 150974 bases at least Q40
            Consensus quality: 151282 bases at least Q30
            Consensus quality: 151487 bases at least Q20
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.

            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *     1    1828: contig of 1828 bp in length
            *  1829    1928: gap of unknown length
            *  1929    5591: contig of 3663 bp in length
            *  5592    5691: gap of unknown length
            *  5692    8607: contig of 2916 bp in length
            *  8608    8707: gap of unknown length
            *  8708   15086: contig of 6379 bp in length
            * 15087   15186: gap of unknown length
            * 15187   28762: contig of 13576 bp in length
            * 28763   28862: gap of unknown length
            * 28863   65227: contig of 36365 bp in length
            * 65228   65327: gap of unknown length
            * 65328  153988: contig of 88661 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153988
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-286G16"
     misc_feature    1. .1828
                     /note="assembly_name:Contig52"
     misc_feature    1929. .5591
                     /note="assembly_name:Contig53"
     misc_feature    5692. .8607
                     /note="assembly_name:Contig54"
     misc_feature    8708. .15086
                     /note="assembly_name:Contig55"
     misc_feature    15187. .28762
                     /note="assembly_name:Contig56"
     misc_feature    28863. .65227
                     /note="assembly_name:Contig57"
     misc_feature    65328. .153988
                     /note="assembly_name:Contig58"
ORIGIN

Query Match       91.1%;    Score 16.4;  DB 2;   Length 153988;
Best Local Similarity   94.4%;   Pred. No. 2.8e+02;
Matches   17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 TTGGTCATCCATGGCTCT 18
          |||| |||||||||||||
Db  57139 TTGGACATCCATGGCTCT 57156

RESULT 47
AC023084
LOCUS       AC023084          157081 bp   DNA    linear   HTG 01-MAR-2000
DEFINITION  Homo sapiens clone RP11-420L4, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC023084
VERSION     AC023084.2  GI:7139796
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157081)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-420L4
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 157081)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,C., Jones,C., Kann,L., Karatas,A., Klein,J.,
```

```
TITLE      Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
           Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
           McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
           Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
           Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
           Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
           Zimmer,A. and Zody,M.
JOURNAL    Direct Submission
           Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Mar 1, 2000 this sequence version replaced gi:6939429.
           All repeats were identified using RepeatMasker:
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           --------------------------  Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           --------------------------  Project Information
           Center project name: L6465
           Center clone name: 420_L4
           --------------------------  Summary Statistics
           Sequencing vector: M13; M77815; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 152870 bases at least Q40
           Consensus quality: 155283 bases at least Q30
           Consensus quality: 155960 bases at least Q20
           Insert size: 156281; sum-of-contigs
           Quality coverage: 5.0 in Q20 bases; sum-of-contigs
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 9 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *      210:  contig of 209 bp in length
         *      309:  gap of 100 bp
         *     5667:  contig of 5358 bp in length
         *     5668:  gap of 100 bp
         *     5767:  contig of 6106 bp in length
         *    11873:  gap of 100 bp
         *    11973:  contig of 10396 bp in length
         *    22369:  gap of 100 bp
         *    22470:  contig of 9269 bp in length
         *    31738:  gap of 100 bp
         *    31839:  contig of 25622 bp in length
         *    57461:  gap of 100 bp
         *    57561:  contig of 26414 bp in length
         *    83975:  gap of 100 bp
         *    84075:  contig of 30550 bp in length
         *   114625:  gap of 100 bp
         *   114725:  contig of 42357 bp in length.
FEATURES             Location/Qualifiers
     source          1..157081
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-420L4"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..209
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    310..5667
                     /note="assembly_fragment"
     misc_feature    5768..11873
```

```
     misc_feature    11974..22369
                     /note="assembly_fragment"
     misc_feature    22470..31738
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    31839..57460
                     /note="assembly_fragment"
     misc_feature    57561..83974
                     /note="assembly_fragment"
     misc_feature    84075..114624
                     /note="assembly_fragment"
     misc_feature    114725..157081
                     /note="assembly_fragment"
ORIGIN

Query Match          91.1%;   Score 16.4;  DB 2;  Length 157081;
Best Local Similarity 94.4%;   Pred. No. 2.8e+02;
Matches   17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        1 TTGGTCATCCATGGCTCT 18
            |||||||||||| ||||||
Db   129259 TTGGTCATCCTTGGCTCT 129276


RESULT 48
AC013643/c
LOCUS       AC013643               157324 bp    DNA     linear   PRI 18-FEB-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-16P20, complete sequence.
ACCESSION   AC013643
VERSION     AC013643.10  GI:18699958
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157324)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-16P20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157324)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 157324)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,R., Lehoczky,J., Levine,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
```

```
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,J., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (28-SEP-2001) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 157324)
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
          Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
          Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
          Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
          Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,J., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
          Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
          Macdonald,P., Major,J., Marquis,R., Matthews,C., McCarthy,M.,
          McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
          Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
          Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
          Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
          Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
          Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
          Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 18, 2002 this sequence version replaced gi:15799620.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
          ---------- Project Information
            Center project name: L3610
            Center clone name: 16_P_20

FEATURES             Location/Qualifiers
     source          1..157324
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-16P20"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(71..230)
                     /rpt_family="MIR3"
     repeat_region   239..292
                     /rpt_family="MIR"
     repeat_region   complement(560..615)
                     /rpt_family="MER5A"
     repeat_region   862..922
                     /rpt_family="G-rich"
     repeat_region   1039..1349
                     /rpt_family="AluSx"
     repeat_region   complement(1631..1796)
                     /rpt_family="LIME"
     repeat_region   1806..1841
                     /rpt_family="AT_rich"
     repeat_region   1857..1994
                     /rpt_family="MIR3"
     repeat_region   2236..2330
                     /rpt_family="MIR"
     repeat_region   3301..3451
                     /rpt_family="MER45A"
     repeat_region   3903..4076
                     /rpt_family="MIR3"
     repeat_region   complement(4656..5130)
                     /rpt_family="L3"
     repeat_region   5550..5855
                     /rpt_family="AluJo"
     repeat_region   5865..5965
                     /rpt_family="L1MA6"
     repeat_region   complement(5979..6505)
                     /rpt_family="L1PA12"
     repeat_region   6507..6673
                     /rpt_family="L1MA6"
     repeat_region   6698..6968
                     /rpt_family="AluSc"
     repeat_region   6972..6994
                     /rpt_family="(CAAA)n"
     repeat_region   6997..7035
                     /rpt_family="L1MA6"
     repeat_region   complement(7567..7740)
                     /rpt_family="MER5B"
     repeat_region   complement(7844..7890)
                     /rpt_family="L3"
     repeat_region   complement(7891..8200)
                     /rpt_family="AluSx"
     repeat_region   complement(8201..8517)
                     /rpt_family="L3"
     repeat_region   8714..8806
                     /rpt_family="CT-rich"
     repeat_region   8906..8942
                     /rpt_family="(TG)n"
     repeat_region   complement(9004..9175)
                     /rpt_family="LIME"
     repeat_region   complement(9285..9462)
                     /rpt_family="LIME"
     repeat_region   9865..9908
                     /rpt_family="(TG)n"
     repeat_region   complement(10140..10441)
                     /rpt_family="AluJb"
     repeat_region   10590..10913
                     /rpt_family="AluJo"
     repeat_region   11103..11134
                     /rpt_family="AT_rich"
     repeat_region   complement(11324..11759)
                     /rpt_family="MSTC"
     repeat_region   complement(11997..12223)
                     /rpt_family="L3"
     repeat_region   13384..13725
                     /rpt_family="MLT1J2"
     repeat_region   14491..14975
                     /rpt_family="MER66B"
     repeat_region   15540..15922
                     /rpt_family="MLT1J2"
     repeat_region   15957..16247
                     /rpt_family="AluSx"
     repeat_region   16248..16630
                     /rpt_family="AluY"
     repeat_region   16744..17233
                     /rpt_family="MLT2D"
     repeat_region   complement(18366..19103)
                     /rpt_family="L1MA4"
     repeat_region   19151..19360
                     /rpt_family="MIR"
     repeat_region   19548..19646
                     /rpt_family="MIR"
     repeat_region   20122..20748
                     /rpt_family="LTR31"
```

```
repeat_region    20751..20782
                 /rpt_family="L2"
repeat_region    20908..21201
                 /rpt_family="AluSx"
repeat_region    complement(21225..21310)
                 /rpt_family="MIR3"
repeat_region    21632..21658
                 /rpt_family="GC_rich"
repeat_region    21732..21811

Query Match      91.1%;   Score 16.4;  DB 9;  Length 157324;
Best Local Similarity 94.4%;  Pred. No. 2.8e+02;
Matches 17;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     1 TTGGTCATCCATGGCTCT 18
         |||||||| |||||||||
Db 11989 TTGGTCATGCATGGCTCT 11972

RESULT 49
AL445306
LOCUS       AL445306    158022 bp   DNA   linear   PRI 05-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-62C22 on chromosome X, complete
            sequence.
ACCESSION   AL445306
VERSION     AL445306.7   GI:11595392
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158022)
AUTHORS     Heath,P.
TITLE       Direct Submission
JOURNAL     Submitted (05-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 6, 2000 this sequence version replaced gi:11545096.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with the source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-62C22 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-62C22 The true
            left end of clone RP5-1074I8 is at 106557 in this sequence.
FEATURES    Location/Qualifiers
source      1..158022
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="X"
            /clone="RP11-62C22"
            /clone_lib=RPCI-11.1"
ORIGIN
```

```
Query Match      91.1%;   Score 16.4;  DB 9;  Length 158022;
Best Local Similarity 94.4%;  Pred. No. 2.8e+02;
Matches 17;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     1 TTGGTCATCCATGGCTCT 18
         || |||||||||||||||
Db  7294 TTTGTCATCCATGGCTCT 7311

RESULT 50
AF440523/c
LOCUS       AF440523    158230 bp   DNA   linear   BCT 12-NOV-2002
DEFINITION  Pseudomonas aeruginosa strain C genomic sequence, gene island
            PAGI-2(C).
ACCESSION   AF440523
VERSION     AF440523.1   GI:24461522
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa
ORGANISM    Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 158230)
AUTHORS     Larbig,K.D., Christmann,A., Johann,A., Klockgether,J., Hartsch,T.,
            Merkl,R., Wiehlmann,L., Fritz,H.J. and Tuemmler,B.
TITLE       Gene Islands Integrated into tRNA(Gly) Genes Confer Genome
            Diversity on a Pseudomonas aeruginosa Clone
JOURNAL     J. Bacteriol. 184 (23), 6665-6680 (2002)
PUBMED      12426355
REFERENCE   2  (bases 1 to 158230)
AUTHORS     Larbig,K.D., Christmann,A., Johann,A., Hartsch,T., Merkl,R.,
            Klockgether,J., Fritz,H.-J. and Tuemmler,B.
TITLE       Direct Submission
JOURNAL     Submitted (30-OCT-2001) Klinische Forschergruppe, Medizinische
            Hochschule Hannover, Carl-Neuberg-Strasse 1, Hannover 30623,
            Germany
FEATURES    Location/Qualifiers
source      1..158230
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="C"
            /db_xref="taxon:287"
            /map="between the genes PA2846 and PA2795"
            /clone="C"
misc_feature 1..27285
            /note="hypervariable genome region"
            /note="corresponding genome region to Pseudomonas
            aeruginosa PAO1 from gene PA2846 to gene PA2820 with an
            average nucleotide substitution rate of 0.35%; all
            annotated genes are conserved and show only a few amino
            acid substitutions"
tRNA        26972..27047
            /product="tRNA-Glu"
tRNA        27050..27125
            /product="tRNA-Gly"
tRNA        27210..27285
            /product="tRNA-Gly"
repeat_region 27270..27285
            /note="att L site; flanks Pseudomonas aeruginosa C gene
            island PAGI-2(C); 16 terminal nucleotides of the 3' end of
            the Gly-tRNA gene"
            /rpt_type=direct
misc_feature 27286..132240
            /note="Pseudomonas aeruginosa C gene island PAGI-2(C),
            integrated into the second Gly-tRNA gene"
repeat_region 27327..27354
            /note="AT-rich repeat"
            /rpt_type=inverted
gene        27514..29445
            /gene="int"
            /note="ORF C1"
CDS         27514..29445
            /gene="int"
```

```
                    /note="bacteriophage P4 integrase subfamily; similar to
                    Xylella fastidiosa XF1718 (int) and Pseudomonas aeruginosa
                    SG17M ORF SG1"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="putative integrase"
                    /protein_id="AAN62094.1"
                    /db_xref="GI:24461523"
                    /translation="MAISDLTVRQAKAAKKTYSIPDTDGLGLVVAPTGGKSWHLRYYW
                    LGKQKRISLGNYPEIGLREARLRDEARALLAQGINPHTDRKQKRHAVKLASDYTFKA
                    VFDAWVEHRAKELKEGRNSTLSQIKRIFEKDVLPSLKQMSIYDIRRPQLLGVLARIEK
                    REAFTTAEKVRTWLGQLFRYALVIVEGMEANPATDLDVVAEPKPAVTHNPYLHLPELP
                    EFLQKLRLYNPRGWQTQLGIRLLFLTGVRTGELRLATPDQFDLDRGLWIIPQIVKQL
                    QDEMRKAGKRPQDVPPYIVPLSLQAIBIVRYILLGVMRPAQRLLTHRSELKKRISENT
                    LNAALKRMGYEDKLTCHGIRGTISTALNEIGYPKIWDAQLSHSDPNKVSSAYNHAKY
                    VEPRRRMMQDWADRLDLLEQGQVEAASAHLTIHIEGVPAMAEEDKPDAIVAASSAPPV
                    PPVVATPIVVTPNEGGITFQRLSQVPSPTHAPEPEVSAIQREREEMLAIYESPSSLP
                    VPLFGKLAGKSKDQINRELKAGKLLSLSLGNRGQRYPDWQLVPLKHKLAQVLMRQYPQ
                    ADSWELYRMLTQPHPDLGADRAAIDIVTPSNLGMVVQIIAGSEPHANAPEVVPPRPISE
                    EVRQSVRRLMESAVALDGA"
gene            complement(29645..30301)
                    /gene="ORF C2"
CDS             complement(29645..30301)
                    /gene="ORF C2"
                    /note="similar to Xylella fastidiosa XF1719"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAN62095.1"
                    /db_xref="GI:24461524"
                    /translation="MHDATSQFPHLRLAIAPGVPSSCFSSLLALQRAEEPEITIAFFE
                    TSSDDLVTGIDEGRYDAGMSLRAVSAPSLKSQPLWIENIAVAMPLWSPLLAQTKLTIA
                    ELLDYPVFRWPAESCSLLDDRLSSLPSLSRQSVQHVTSFDMLALWVGAGYGVGITAQS
                    RIERADAWGITARPLAEGPYEVVTHLQRLKGRTNAVSERFERRAMQVAKDSAVQSNTR
                    "
gene            30421..30714
                    /gene="ORF C3"
CDS             30421..30714
                    /gene="ORF C3"
                    /note="similar to Xylella fastidiosa XF1720"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAN62096.1"
                    /db_xref="GI:24461525"
                    /translation="MQKSTIQPGRPSGTTTYESAAATAFGQAVRAARVAQGVAQDEFA
                    SRAGISRSHMGKIERGEHVPTLPLILKISVALGISAAELMAATERNLRAETDL"
gene            complement(30736..31626)
                    /gene="bphR"
                    /note="ORF C4"
CDS             complement(30736..31626)
                    /gene="bphR"
                    /note="similar to Pseudomonas aeruginosa SG17M ORF SG105"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="putative transcriptional regulator"
                    /protein_id="AAN62138.1"
                    /db_xref="GI:24461567"
                    /translation="MELRHLRYFMALAEELHFTRAAERLHIEQPPLSRAIKELEDELG
                    VVLFDRNRRGTVLTSAGVVFLQDIRRLFTVLEQARENAQAVAAGLRGSLRAISDGAI
                    DARLSAFLARCRAEPEIEIRLSEVPLAEQLRGLRSDFMIGFAHTADVGGDIRAEPI
                    WRDPLVIAVPARHALLAHKKVPLQEVGIHPLVLCDPHACEGYCRELTRLLQVLEHKLN
                    VVEEVSLDMMLTIVGAGYGIGFLAASKIPVCQRPDVVIRPLAIDSAVITTYLLRADS
                    SDLSASLERFIVRLRDSASG"
gene            complement(31664..31987)
                    /gene="ORF C5"
CDS             complement(31664..31987)
                    /gene="ORF C5"
                    /codon_start=1
```

```
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAN62097.1"
                    /db_xref="GI:24461526"
                    /translation="MTQDTQRPTLKVILHAPTAEAFERARSNATNLKRAAPADVRIV
                    ANAQAVAAALDTAPSDLDALTWVCPNTLSRIGRDNREPLEVLDGPAVIEMARMQQAGW
                    IYIRA"
gene            complement(32019..33395)
                    /gene="ORF C6"
CDS             complement(32019..33395)
                    /gene="ORF C6"
                    /codon_start=1
                    /transl_table=11
                    /product="putative pyridine nucleotide-disulfide
                    oxidoreductase, class I"
                    /protein_id="AAN62098.1"
                    /db_xref="GI:24461527"
                    /translation="MQTRKVDVAIIGAGTAGSTAFHALKATGKQVVLIDRGPLGTTCA
                    RVGCMGPSKAAIHAGMHWKVAKALAASATPSTGQPQAIMSHARQTRDALAGAAAERTRK
                    AAGESLIMAEARFVAPGILQAGEQRIEASAFIVATGSRPVVPKALDGVASRILTTDTL
                    FDLGVLPARMGILGLGAIGLELGLALSRLDVQVVAADQKDAIGGIQDPDVLERAIARF
                    ETELPMWLGAAVEVSLEGDQVRMRAGERDALVDRLLVVTGRQPNTEALDLAAAGISLD
                    QAGRPSIDPATMQACGPSQVPIFFAGDVQPDRPLMHEAADEGQMAAQAALASLRGRSW
                    PGASRRVPITILFTDPDACAVGMTYRAAVDEGAVVGTAEGSGNGRSKILGAPENLLHI
                    YADPGSGALLGASMLJTQGEHLAHLIAWAIQAKQTVNDLLAMPYYHPSIEEMLQSALK
                    SASQQMRP"
gene            complement(33438..34244)
                    /gene="ORF C7"
CDS             complement(33438..34244)
                    /gene="ORF C7"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="conserved hypothetical membrane protein"
                    /protein_id="AAN62099.1"
                    /db_xref="GI:24461528"
                    /translation="MLSLTGGVLGLIIGAVLGLTGAGGGIFAVPALVFGLGMDIRQAA
                    FVALLAVGAAATLGALQGLRQGVVRYKAAMMLAAAGAVTAPLGVQFAHWLSPRWLNLI
                    FVAIMLVAYRMFMSSRGSCTQDDLIADEPARVCKISKDTGRFVWNVRTATTLGSIGIV
                    SGLATGMLGVGGGFIIVPALAHFSELRMHSIVATSLMVIALLSAVTVFIAWSHGMTLT
                    APAWTFVLTALVGMSLGRVLARRIPSKMLQRVFSITCVAVAALMRNVG"
gene            complement(34335..35108)
                    /gene="dsbG"
                    /note="ORF C8"
CDS             complement(34335..35108)
                    /gene="dsbG"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="putative thiol:disulfide interchange protein"
                    /protein_id="AAN62100.1"
                    /db_xref="GI:24461529"
                    /translation="MFAQRSHISFLLAASFMVVTGCSKAETADKPAVLKALESGQGLTV
                    TQEFKVGGLRAFAAVAGDRPIAVYITSDGNAIVGTRLNAKGBPMDEADLEKLAAKPV
                    SDKEWSQLQSSTWIDGKADAPRVIYTFSDANCPYCNAFWEARAPWVDSGKVQLRHIL
                    VGIIKDDSPAKAAILGAPDRSAALLNERQFGQGGITPAKSVPADVRKILLDDNLALM
                    ASTGFRGTPGIVVRGPNGILKKYNGMEPRGAQLGEVLGPR"
gene            complement(35111..35947)
                    /gene="ORF C9"
CDS             complement(35111..35947)
                    /gene="ORF C9"
                    /codon_start=1
                    /evidence=not_experimental
                    /transl_table=11
                    /product="putative thiol:disulfide interchange protein"
                    /protein_id="AAN62101.1"
```

```
Query Match                    91.1%;   Score 16.4;   DB 1;   Length 158230;
Best Local Similarity   94.4%;   Pred. No. 2.8e+02;
Matches   17;   Conservative    0;   Mismatches   1;   Indels    0;   Gaps    0;

Qy        1 TTGGTCATCCATGGCTCT 18
            ||||||| |||||||
```

us-10-054-387-40.rge

Db   82738 TTGGTCATCCGTGGCTCT 82721

Search completed: June 7, 2004, 16:24:18
Job time : 1581 secs